# AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY



PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXIII, Number 4
OCTOBER, 1941

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

# CONTENTS



The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is \$7.50 (foreign postage 50 cents); single numbers \$2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications may be sent to Professor F. D. MURNAGHAN at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

o
re
𝔇
fo
o
m
ec
N
d

ia
o
re
p
w
w

w
in
ex
el
in

(
vo

A
by
§§
be

# COMPLETE DIFFERENCE IDEALS.*

By J. F. Ritt.

If the manifold $\mathfrak{M}$ of an ideal $\Sigma$ of differential polynomials is composed of two manifolds $\mathfrak{M}_1$ and $\mathfrak{M}_2$ with no solution in common, $\Sigma$ has a unique representation as the product of two ideals whose manifolds are respectively $\mathfrak{M}_1$ and $\mathfrak{M}_2$.[1] We show here that no result of this degree of generality holds for difference polynomials,[2] and then direct our attention to a natural class of difference ideals, the *complete* ideals, for which a theorem on separated manifolds can be derived. While we are chiefly interested here in difference equations, our decomposition theorem, Theorem II of § 10, is distinctly Noetherian in form, so general, even, as to unite the theory of algebraic difference equations with the theory of numbers.

## Difference Ideals and their Products.

**1.** We use a difference ring $\mathcal{R}$ as in R. R., § 1. In defining *difference ideal* in R. R., § 3, it was stipulated that the presence in a difference ideal $\pi$ of the transform $a_1$ of an element $a$ must imply the presence in $\pi$ of $a$. This restriction upon the nature of a difference ideal, made to insure that every prime ideal be perfect, is a potential source of inelegance for the investigation which we are about to undertake. In meeting this situation, our first step will be to call the difference ideals of R. R. *reflexive difference ideals.*

In the present paper, an ideal $\pi$ (in the sense of algebra) contained in $\mathcal{R}$ will be called a *difference ideal* if the presence in $\pi$ of the element $a$ of $\mathcal{R}$ implies the presence in $\pi$ of the transform of $a$. A difference ideal $\pi$ can be extended into a reflexive difference ideal by the adjunction to $\pi$ of every element which has a transform of some order in $\pi$; one obtains thus the intersection of all reflexive ideals containing $\pi$.

* Received February 17, 1941.

[1] Ritt, *Proceedings of the National Academy of Sciences of the U. S. A.*, vol. 25 (1939), p. 90; Ritt and Kolchin, *Bulletin of the American Mathematical Society*, vol. 45 (1939), p. 895.

[2] Difference ideals are treated by Ritt and Raudenbush, *Transactions of the American Mathematical Society*, vol. 46 (1939), p. 445. This paper will be denoted by R. R. We would mention here that minor modifications of the argument in R. R., §§ 11, 12 lead to the result that every infinite system $\Sigma$ of difference polynomials can be derived from any of its bases with a finite number of shufflings.

A difference ideal $\pi$ will be called *perfect* if whenever $a$ is such that some product of integral powers of transforms of $a$ is in $\pi$, $a$ is also contained in $\pi$. A perfect ideal as thus defined is reflexive; hence the perfect ideals of the present paper are the same ideals which were called perfect in R. R.

A difference ideal $\pi$ will be called *prime* if $\pi$ is a prime ideal in the sense of algebra; a prime ideal is perfect only if it is reflexive.

For a ring $\mathfrak{R}$ with a basis theorem, the theorem of R. R., § 8 now becomes: *Every perfect ideal in $\mathfrak{R}$ has a unique representation as the intersection of a finite number of prime ideals none of which contains any other. In this representation, the prime ideals are perfect.*

**2.** Given a set $\sigma$ of elements of $\mathfrak{R}$, we denote by $[\sigma]$ the difference ideal[3] generated by $\sigma$, that is, the intersection of all ideals containing $\sigma$. An element $a$ is in $[\sigma]$ if and only if $a$ is a linear combination, with coefficients in $\mathfrak{R}$, of elements of $\sigma$ and their transforms of various orders. We denote by $\{\sigma\}$ the perfect ideal generated by $\sigma$.

The *product* $\sigma\tau$ of two ideals $\sigma$ and $\tau$ will be defined as the ideal generated by the totality of elements $ab$ where $a$ is any element in $\sigma$ and $b$ any in $\tau$. The product, as thus defined, of $\sigma$ and $\tau$ is identical with their product in the usual sense of algebra; this results from the fact that the transform of the product of two elements is the product of their transforms.

Let $\sigma$ and $\tau$ be ideals. We shall prove that

$$\{\sigma\tau\} = \{\sigma\} \cap \{\tau\}. \tag{1}$$

The second member of (1) contains the first member. We show that the first member contains the second, thereby establishing not only (1), but also the relation

$$\{\sigma \cap \tau\} = \{\sigma\} \cap \{\tau\}. \tag{2}$$

In the lemma which follows, the meaning of $\sigma_n$, for any set $\sigma$, will be as in R. R., § 5.[4]

LEMMA. *Let $\sigma$ and $\tau$ be ideals. For $n \geqq 1$, if $a \epsilon \sigma_n$ and $b \epsilon \tau_n$, $ab \epsilon (\sigma\tau)_n$.*[5]

First, let $n = 1$. Some product $c$ of transforms[6] of $a$ is in $\sigma$. Some product $d$ of transforms of $b$ is in $\tau$. Then $cd \epsilon \sigma\tau$. Some product of trans-

[3] In the remainder of this paper, *ideal* will mean *difference ideal.*

[4] One will notice that square brackets as used there have a meaning identical with that established here.

[5] The parentheses about $\sigma\tau$ are merely symbols of inclusion.

[6] More fully, a product of powers of $a$ and its transforms.

forms of $ab$ is a multiple of $cd$. Then $ab \epsilon (\sigma\tau)_1$. Now let $n = 2$. There are a $c$ and a $d$, related as above to $a$ and to $b$, with expressions

$$c = g_1u_1 + \cdots + g_ru_r, \qquad d = h_1v_1 + \cdots + h_sv_s$$

where the $u$ are in $\sigma_1$ and the $v$ in $\tau_1$. Now $cd$ is a sum of terms $ku_iv_j$. By the case of $n = 1$, each of these terms is in $(\sigma\tau)_1$. This puts $ab$ in $(\sigma\tau)_2$. The proof continues by induction.

We now complete the proof of (1). Let $a$ be in $\{\sigma\}$ and in $\{\tau\}$. There is an $n$ such that $a$ is in $\sigma_n$ and in $\tau_n$. By the lemma, $a^2$ is in $(\sigma\tau)_n$ so that $a$ is in $(\sigma\tau)_n$ and thus in $\{\sigma\tau\}$.

We have, immediately, the theorem:

THEOREM I. *For any ideals* $\sigma_1, \cdots, \sigma_s$,

$$\{\sigma_1\sigma_2\cdots\sigma_s\} = \{\sigma_1 \cap \cdots \cap \sigma_s\} = \{\sigma_1\} \cap \cdots \cap \{\sigma_s\}.$$

## Abstract Manifolds.

**3.** It is well to make clear at this point the notion of the manifold of a system of difference polynomials (forms). Let $\mathcal{F}$ be a difference field and $\Sigma$ a system of forms in the unknowns $y_1, \cdots, y_n$ with coefficients in $\mathcal{F}$. A difference field $\mathcal{F}'$ is called an extension of $\mathcal{F}$ if $\mathcal{F}'$ contains $\mathcal{F}$. Here it is understood that if $a$ is an element of $\mathcal{F}$ and if $a'$ is the transform of $a$ before $\mathcal{F}$ is absorbed into $\mathcal{F}'$, $a'$ remains the transform of $a$ after the absorption. If an extension of $\mathcal{F}$ contains a set of $n$ elements $a_1, \cdots, a_n$ which are such that the substitution for each $y_i$ and its transforms, in the forms of $\Sigma$, of $a_i$ and its corresponding transforms reduces each form of $\Sigma$ to 0, we shall call $a_1, \cdots, a_n$ a solution of $\Sigma$. The totality of solutions of $\Sigma$, obtained from all extensions of $\mathcal{F}$ in which solutions exist, will be called the *manifold* of $\Sigma$.[7]

A situation is met in connection with difference equations which is anomalous insofar as it does not occur in the theory of algebraic equations. It is possible for each of two systems $\Sigma_1$ and $\Sigma_2$ with coefficients in a single difference field $\mathcal{F}$ to have a non-vacuous manifold and still for no extension of $\mathcal{F}$ to exist in which $\Sigma_1$ and $\Sigma_2$ both have solutions.[8] Let $\mathcal{F}$ consist of all rational numbers. We use forms in the single unknown $y$. Let $\Sigma_1$ consist of

---

[7] Here we follow a standard procedure of algebra, inaugurated by van der Waerden. Vague in nature as the totality of solutions may be, considerations of language justify its toleration; in any particular situation, there is no vagueness.

[8] It is to be emphasized that we are talking of solutions from the standpoint of algebra and not from that of analysis.

$y^2+1$ and $y_1-y$;[9] $\Sigma_2$ of $y^2+1$ and $y_1+y$. $\Sigma_1$ has the solution $y=(-1)^{\frac{1}{2}}$ in the field of complex numbers, each number its own transform; $\Sigma_2$ has the solution $y=(-1)^{\frac{1}{2}}$ in the field of complex numbers with the transform of each number taken as the conjugate of the number. Suppose that there is an $\mathfrak{F}'$ in which $\Sigma_1$ and $\Sigma_2$ both have solutions. Let $a$ be a solution of $\Sigma_1$ in $\mathfrak{F}'$. Then $-a$ is the only other solution of $\Sigma_1$ in $\mathfrak{F}'$. Now $a$ and $-a$ are the only possibilities for solutions of $\Sigma_2$ in $\mathfrak{F}'$. However, as they annul $y_1-y$ and are not zero, they do not annul $y_1+y$.

## Separated Ideals.

**4.** Two ideals $\sigma$ and $\tau$ in $\mathfrak{R}$ will be said to be *separated* if the perfect ideal generated by the logical sum of $\sigma$ and $\tau$ contains unity; that is, in symbols, if $1 \epsilon \{\sigma, \tau\}$. If $\sigma$ and $\tau$ are ideals of forms, they are separated if and only if they have no common solution.[10]

Two ideals $\sigma$ and $\tau$ will be said to be strongly separated if elements $a$ and $b$, belonging respectively to $\sigma$ and to $\tau$, exist such that $a+b=1$.

We give an example of two perfect ideals of forms in $y$ which are separated but not strongly separated. We use the field of rational numbers.

Let

$$A=1+yy_1; \quad B=y+y_1; \quad \Sigma=\{A,B\}.$$

As

$$1-y^2=A-yB,$$

$\Sigma$ contains $1-y^2$. Let $C=(1-y)(1+y_1)$. The product of $C$ by its transform is divisible by the transform of $1-y^2$. Hence $C \epsilon \Sigma$. If we replace $y_1$ in $C$ by $B-y$, we find that $(1-y)^2 \epsilon \Sigma$. The substitution $y=B-y_1$ into $C$ shows that $(1+y_1)^2 \epsilon \Sigma$. Then $(1+y)^2 \epsilon \Sigma$. Thus $\Sigma$ contains $1-y$ and $1+y$ and hence contains unity. Hence $\{A\}$ and $\{B\}$ are separated.

Now if $A=0$ and $B=0$ are regarded as difference equations in the analytic function $y$ of the variable $x$, $B=0$ is satisfied by $e^{\pi i x}$ and $A=0$ is satisfied by $-ie^{g(x)}$ where $g(x)=(\pi i/2)e^{\pi i x}$. Each of these analytic functions assumes the value 1 for even integral values of $x$ and the value $-1$ for odd values of $x$. Thus all forms in $\{A\}$ and in $\{B\}$ vanish when each $y_j$ with $j$ even is replaced by 1 and each $y_j$ with $j$ odd by $-1$. Hence $\{A\}$ and $\{B\}$ are not strongly separated.

The strong separation of two ideals of forms $\Sigma_1$ and $\Sigma_2$ thus involves the

---

[9] $y_i=y(x+i)$.

[10] This follows from the *Nullstellensatz* of R. R., § 13.

incompatibility of the systems $\Sigma_1 = 0$ and $\Sigma_2 = 0$ considered as systems of algebraic equations in an infinite set of unknowns.

Strongly separated ideals are met in connection with algebraically irreducible forms whose manifolds are reducible. Thus, let

$$A = (y_1 - y)^2 - 2(y_1 + y) + 1.$$

Then

(3) $$\{A\} = \{A, B\} \cap \{A, C\}$$

where

$$B = y_2 - y, \qquad C = y_2 - 2y_1 + y - 2.$$

Unity is a linear combination, with constant coefficients, of $B$, $C$ and their first transforms. Thus the ideals in the second member of (3) are strongly separated.

**5.** If $\sigma$ and $\tau$ are strongly separated ideals, $\sigma\tau = \sigma \cap \tau$. For the proof let $a + b = 1$ with $a$ in $\sigma$ and $b$ in $\tau$. Let $c \epsilon \sigma \cap \tau$. As $c = ca + cb$, $c \epsilon \sigma\tau$.

If $\sigma_1, \cdots, \sigma_s$ are ideals with $\sigma_1$ strongly separated from each of $\sigma_2, \cdots, \sigma_s$, then $\sigma_1$ is strongly separated from $\sigma_2\sigma_3 \cdots \sigma_s$. For, let $a_i + b_i = 1$, $i = 2, \cdots, s$, with $a_i$ in $\sigma_1$ and $b_i$ in $\sigma_i$. Then

$$b_2 \cdots b_s = (1 - a_2) \cdots (1 - a_s) = 1 - c$$

and here $b_2 \cdots b_s \epsilon \sigma_2 \cdots \sigma_s$ while $c \epsilon \sigma_1$. *A fortiori*, $\sigma_1$ is strongly separated from $\sigma_2 \cap \cdots \cap \sigma_s$.

**6.** A set of ideals $\sigma_1, \cdots, \sigma_s$ will be said to be strongly separated *in pairs* if each $\sigma_i$ is strongly separated from each $\sigma_j$ with $j \neq i$. For such ideals, by § 5,

$$\sigma_1\sigma_2 \cdots \sigma_s = \sigma_1 \cap \cdots \cap \sigma_s.$$

## An Indecomposable Ideal.

**7.** We use the field of rational numbers. Let $A = y^2 - 1$, $\Sigma = [A]$. The forms of $\Sigma$ are of the type

(4) $$MA + M'A_1 + \cdots + M^{(r)}A_r$$

where $A_j$ is the $j$-th transform of $A$.

The manifold of $\Sigma$ consists of 1 and $-1$. We proceed to show that $\Sigma$ has no representation either as a product or as an intersection of two ideals whose manifolds consist respectively of 1 and $-1$.

Let $\Sigma_1$ and $\Sigma_2$ be ideals with the respective manifolds 1 and $-1$. Then $\Sigma_1$ contains a form $B$ which does not vanish when each $y_j$ is replaced by $-1$ and $\Sigma_2$ contains a form $C$ which does not vanish when each $y_j$ is replaced by 1.

If $k$ is sufficiently large, the $k$-th transform $C_k$ of $C$ will contain no $y_j$ which appears in $B$; we choose such a $k$. Then $BC_k$ fails to vanish when the $y_j$ in $B$ are replaced by $-1$ and those in $C_k$ by 1. Now if $\Sigma$ were the product or the intersection of $\Sigma_1$ and $\Sigma_2$, $\Sigma$ would contain $BC_k$ and $BC_k$ would be of type (4). Then $BC_k$ would have to vanish when the $y_j$ are replaced in any manner, independently of one another, by 1 or $-1$. This proves the non-existence, for $\Sigma$, of the described representations.

The system of equations $\Sigma = 0$, considered as an algebraic system in an infinite set of unknowns $y_j$, has, in addition to the solutions $y_j = 1$ and $y_j = -1$, $j = 0, 1, \cdots$, an abundance of other solutions, which cannot be used for the construction of analytic solutions of the system $\Sigma$ of difference polynomials. One may, on this basis, properly blame the algebraic system for the indecomposability of $\Sigma$. The thought now arrives that an ideal $\Lambda$ will be decomposable if $\Lambda$ contains some power of every form in $\{\Lambda\}$. The restriction suggested here is actually too strong. It suffices that $\Lambda$ contain, for every form $G$ in $\{\Lambda\}$, some power of some transform of $G$. Speaking in the ignorance which exists with respect to the analytic solutions of non-linear difference equations, one might make the vague statement that it is permissible for the algebraic system $\Lambda = 0$ to have some, but not too many, solutions which do not yield analytic solutions of $\Lambda$.

## Complete Ideals.

**8.** An ideal $\sigma$ in $\mathfrak{R}$ will be called *complete* if, for every element $a$ in $\{\sigma\}$, there exist a positive integer $p$ and a non-negative integer $j$, each depending on $a$, such that $a_j$ being the $j$-th transform[11] of $a$, $a_j^p \epsilon \sigma$.

We proceed to prove that, *for an ideal $\sigma$ to be complete, it is necessary and sufficient that the presence in $\sigma$ of a product of transforms of an element should imply the presence in $\sigma$ of a power of a transform of the element.*

The necessity is obvious. For the sufficiency, we consider an ideal $\sigma$ which satisfies the condition, and use the sets $\sigma_n$ of § 2. If $a \epsilon \sigma_1$, some product of transforms of $a$ is in $\sigma$, so that some power of some transform of $a$ is in $\sigma$. Let $a \epsilon \sigma_2$. There is a product $b$ of transforms of $a$ such that

$$b = hu + \cdots + kv \tag{5}$$

with $u, \cdots, v$ in $\sigma_1$. For every $j$ and $p$

$$b_j^p = (h_j u_j + \cdots + k_j v_j)^p.$$

[11] Henceforth, in this paper, transforms will be indicated by subscripts.

By the case of $n = 1$, $b_j^p \epsilon \sigma$ if $j$ and $p$ are both large. Then some power of some transform of $a$ is in $\sigma$. Continuing, we find $\sigma$ to be complete.

The intersection of any *finite* set of complete ideals is complete.

**9.** An ideal $\sigma$ will be called *mixed* if the presence in $\sigma$ of $ab$ implies the presence in $\sigma$ of $ab_1$. A mixed ideal satisfies the italicized condition of § 8. Hence, *every mixed ideal is complete.*

The intersection of any set, finite or infinite, of mixed ideals, is mixed.

**10.** We prove the theorem:

THEOREM II. *Let $\sigma$ be a complete ideal. Suppose that $\{\sigma\}$ is the intersection of $s$ perfect ideals $\tau_1, \cdots, \tau_s$ which are strongly separated in pairs. Then $\sigma$ has one and only one representation as the intersection of $s$ complete ideals $\sigma_1, \cdots, \sigma_s$ with $\{\sigma_i\} = \tau_i$, $i = 1, \cdots, s$. In this representation, $\sigma_1, \cdots, \sigma_s$ are strongly separated in pairs.*

By § 6, the intersection of the $\sigma_i$ equals their product.

**11.** We let $\tau = \{\sigma\}$ and consider first the case of $s = 2$.

Two elements will be said to be congruent, modulo an ideal, if their difference is in the ideal.

Let $e + f = 1$ with $e$ in $\tau_1$ and $f$ in $\tau_2$. Then $ef \epsilon \tau$ so that some $(e_j f_j)^p$ is in $\sigma$. We write

$$e_j + f_j = 1 \tag{6}$$

and raise both sides of (6) to the $(2p-1)$-th power. In the first member of the resulting equation, let $c$ represent the sum of those terms in which the exponent of $e_j$ is not less than $p$ and let $d$ represent the sum of the remaining terms. Then $c \epsilon \tau_1$, $d \epsilon \tau_2$, and

$$c + d = 1 \tag{7}$$

while

$$cd \equiv 0, [\sigma].^{12} \tag{8}$$

By (8), $c_1 d_1 \epsilon \sigma$ so that $(cd_1)(c_1 d_2) \epsilon \sigma$. Thus $cd_1 \epsilon \tau$. Similarly $c_1 d \epsilon \tau$. Thus, for some $j$ and $p$,

$$c^p{}_j d^p{}_{j+1} \equiv 0, [\sigma]; \qquad c^p{}_{j+1} d^p{}_j \equiv 0, [\sigma]. \tag{9}$$

Putting $d = 1 - c$ in (8), we find that $c \equiv c^2, [\sigma]$. This implies that $c \equiv c^q, [\sigma]$ for every $q$. Similar congruences hold for $d$ and for the transforms of $c$ and $d$. Thus, by (9),

$$c_j d_{j+1} \equiv c_{j+1} d_j \equiv 0, [\sigma]. \tag{10}$$

[12] Of course, $[\sigma] = \sigma$.

We now write $a = c_j$, $b = d_j$. Then $a \in \tau_1$, $b \in \tau_2$ and

(11) $$a + b = 1,$$

while

(12) $$ab \equiv ab_1 \equiv a_1 b \equiv 0, [\sigma].$$

**12.** We show now that

(13) $$a_1 - a \equiv b_1 - b \equiv 0, [\sigma].$$

By (12)

(14) $$a(b - b_1) \equiv 0, [\sigma].$$

In (14), we put $a = 1 - b$ and $b - b_1 = a_1 - a$. We find that

(15) $$(1 - b)(a_1 - a) \equiv 0, [\sigma].$$

By (12),

(16) $$a_1 - a \equiv 0, [\sigma].$$

Similarly, $b_1 - b \equiv 0, [\sigma]$.

**13.** We now let $\sigma_1 = [\sigma, a]$ and $\sigma_2 = [\sigma, b]$. We wish to prove that $\{\sigma_1\} = \tau_1$. It is only necessary to show that $\tau_1$ is contained in $\{\sigma_1\}$. Let $g$ be any element in $\tau_1$. Then $bg$ is in $\tau$, hence in $\{\sigma_1\}$. Now $bg = g - ag$. Because $ag$ is in $\sigma_1$, $g$ is in $\{\sigma_1\}$. Similarly $\{\sigma_2\} = \tau_2$.

By (11), $\sigma_1$ and $\sigma_2$ are strongly separated.

**14.** Let us prove that $\sigma_1$ is complete. If $g$ is in $\tau_1$, $gb$ is in $\tau$. Hence some $g_j{}^p b_j{}^p$ is in $\sigma$. As $b_j = 1 - a_j$ and $g_j{}^p a_j$ is in $\sigma_1$, $g_j{}^p$ is in $\sigma_1$. In the same way, $\sigma_2$ is complete.

**15.** We prove now that

(17) $$\sigma = \sigma_1 \cap \sigma_2.$$

It is only necessary to show that the first member of (17) contains the second.

Let $g \in \sigma_1 \cap \sigma_2$. Because $g \in \sigma_1$,

(18) $$g = e + fa + f'a_1 + \cdots + f^{(r)}a_r$$

with $e$ in $\sigma$. By (16), $g$ has an expression

(19) $$g = h + ka$$

with $h$ in $\sigma$. Then $bg \in \sigma$. Similarly $ag \in \sigma$, so that $g$, which equals $(a + b)g$, is in $\sigma$.

**16.** We treat now the question of uniqueness. Let $\sigma$ be the intersection of complete ideals $\sigma'_1$ and $\sigma'_2$ with $\tau_1 = \{\sigma'_1\}$, $\tau_2 = \{\sigma'_2\}$. Let $g \in \sigma'_1$. Some $b_j{}^p$

is in $\sigma'_2$. Thus $b_j{}^p g$ is in $\sigma$ and hence in $\sigma_1$. As $b_j = 1 - a_j$ and $a_j$ is in $\sigma_1$, $g$ is in $\sigma_1$. Again, let $h \epsilon \sigma_1$. Then for every $j$, $hb_j = h(1 - a_j)$ is in $\sigma$ and hence in $\sigma'_1$. It follows that $h(1 - a_j{}^q)$ is in $\sigma'_1$ for every $j$ and $q$. For $j$ and $q$ large, $a_j{}^q \epsilon \sigma'_1$. Thus $h \epsilon \sigma'_1$. We have proved that $\sigma_1$ and $\sigma'_1$ are identical; so also are $\sigma_2$ and $\sigma'_2$.

This settles the case of $s = 2$.

**17.** We treat the case of a general $s$. Let the theorem be true for $s \leqq r - 1$, where $r > 2$. For $s = r$, if we let $\xi = \tau_2 \cap \cdots \cap \tau_r$, we have $\tau = \tau_1 \cap \xi$. By § 5, $\tau_1$ and $\xi$ are strongly separated. Hence $\sigma = \sigma_1 \cap \zeta$ with $\sigma_1$ and $\zeta$ complete and with $\tau_1 = \{\sigma_1\}$, $\xi = \{\zeta\}$. From the case of $r - 1$, we know that $\zeta = \sigma_2 \cap \cdots \cap \sigma_r$ with each $\sigma_i$ complete and $\tau_i = \{\sigma_i\}$. Then $\sigma = \sigma_1 \cap \sigma_2 \cdots \cap \sigma_r$. Because $\sigma_1$ and $\zeta$ are strongly separated and $\zeta$ is contained in each $\sigma_i$ with $i > 1$, $\sigma_1$ is strongly separated from each of $\sigma_2, \cdots, \sigma_r$. In other words, $\sigma_1, \sigma_2, \cdots, \sigma_r$ are strongly separated in pairs.

For the uniqueness, let $\sigma$ have a second representation $\sigma = \sigma'_1 \cap \cdots \cap \sigma'_r$. By Theorem I, $\{\sigma'_2 \cap \cdots \cap \sigma'_r\} = \xi$ where $\xi$ is as above. This, by the case of $s = 2$, means that $\sigma'_1 = \sigma_1$ and that $\sigma'_2 \cap \cdots \cap \sigma'_r$, which is complete, equals $\zeta$. By the case of $s = r - 1$, $\sigma'_i = \sigma_i$, $i = 2, \cdots, r$.

**18.** Let us prove that if $\sigma$, in Theorem II, is mixed, each $\sigma_i$ is mixed. We consider $\sigma_1$. Let $gh \epsilon \sigma_1$. We use $\zeta$ as in § 17. Let $a + b = 1$ where $a \epsilon \sigma_1$, $b \epsilon \zeta$. Then $ghb \epsilon \sigma$, so that $gh_1 b \epsilon \sigma$. As $b = 1 - a$ and $gh_1 a$ is in $\sigma_1$, $gh_1$ is in $\sigma_1$. Similarly, $\sigma_2, \cdots, \sigma_s$ are mixed.

Again if $\sigma$ is reflexive, every $\sigma_i$ is reflexive. Let some transform $c_j$ of an element $c$ be in $\sigma_1$. Then, for $b$ as above, $c_j b_j \epsilon \sigma$ so that $cb$ is in $\sigma$ and thus in $\sigma_1$. As above, $c \epsilon \sigma_1$. Similarly, $\sigma_2, \cdots, \sigma_s$ are reflexive.

**19.** We note that the elements of each $\sigma_i$ in Theorem II are of the type $e + fa^{(i)}$ with $e$ in $\sigma$ and $a^{(i)}$ a fixed element of $\tau_i$.

## Construction of Complete Ideals.

**20.** Given a set $\sigma$ of elements of $\mathcal{R}$, there are complete ideals, for instance $\{\sigma\}$, which contain $\sigma$. There is not always a smallest complete ideal containing $\sigma$; that is, the intersection of all complete ideals containing $\sigma$ may not be complete.

For an example, we consider forms in $y$ over the field of rational numbers.

Let $\Sigma = [yy_1]$. Then $\Sigma$, whose forms are linear in transforms of $yy_1$, is not complete.[13] For any positive integer $p$, let

$$\Sigma_p = [y^p, yy_1].$$

$\Sigma_p$ is complete because $\{\Sigma_p\} = \{y\}$ and every linear combination of transforms of $y^p$ is in $\Sigma_p$. The intersection of all $\Sigma_p$, $p = 1, 2, \cdots$, is $[yy_1]$. To see this, let, for some $p$, $A$ be a form in $\Sigma_p$ which is not in $[yy_1]$. Then $A = B + C$ with $B$ in $[yy_1]$ and $C$ in $[y^p]$. Transferring terms of $C$ to $B$ if necessary, we shall suppose that no term in $C$ is in $[yy_1]$. Let $q$ be an integer which exceeds the total degree of $C$ in the $y_j$. If $A$ were in $\Sigma_q$, $C$ would be in $\Sigma_q$. Writing $C$ as a sum of homogeneous polynomials $H_i$, we would find each $H_i$ to be in $\Sigma_q$ and therefore in $[yy_1]$. This is impossible. There is thus no smallest complete ideal containing $[yy_1]$.

Given any set $\sigma$ of elements of $\mathfrak{R}$, the intersection of all mixed ideals which contain $\sigma$ is a mixed ideal.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

[13] $y \epsilon \{\Sigma\}$ and no $y_j{}^p$ is in $\Sigma$.

# DECOMPOSITIONS OF COMPACT METRIC SPACES.*

By R. L. Wilder.

Let $M$ be a compact metric space and $\psi$ a local topological property. If $\psi$ fails to hold at $x \in M$, we call $x$ a *$\psi$-singular point* of $M$. Let $\Gamma$ denote a class of compact metric spaces having the property that if $M \in \Gamma$ and if $S$ is the set of $\psi$-singular points of $M$, then $\bar{S}$ is either vacuous or contains non-degenerate connected sets; we then call $\psi$ *expansive relative to* $\Gamma$.

We commence with a property $\psi$ that is expansive relative to a class $\Gamma$ of compact metric spaces, and define [4, p. 224] for each $M \in \Gamma$ a new distance function by means of which there is generated a continuous single-valued mapping of $M$ into a new space $M'$: For $x, y \in M$, consider all possible "$\epsilon$-chains" $t_0$ $(=x)$, $t_1, \cdots, t_n$ $(=y)$, where $t_i \in S (0 < i < n$, $S =$ set of $\psi$-singular points of $M$) and $\rho(t_i, t_{i+1}) < \epsilon$ $(0 \leqq i \leqq n-1)$. The greatest lower bound of numbers $\epsilon$ for which such $\epsilon$-chains exist we call the new distance $\rho'(x, y)$. Since $\rho'(x, y) \leqq \rho(x, y)$, the mapping so generated is continuous. If we call each point of $M - \bar{S}$ a *$\psi$-prime part* and each component of $\bar{S}$ a $\psi$-prime part, then the counter images of the points of $M'$ are the $\psi$-prime parts of $M$, and we therefore call $M'$ the space of $\psi$-prime parts of $M$.

Theorem 1. *If the local topological property $\psi$ is expansive relative to a class $\Gamma$ of compact metric spaces and $M \in \Gamma$, then the space $M'$ of $\psi$-prime parts of $M$, if itself an element of $\Gamma$, has no $\psi$-singular points.*[1]

*Proof.* Suppose the set $S'$ of $\psi$-singular points of $M'$ is non-vacuous. Since $M' \in \Gamma$ and $\psi$ is expansive relative to $\Gamma$, $\bar{S}'$ contains a non-degenerate component $C'$.[2] The counter-image $C$ of $C'$ is a subcontinuum of $M$ [4, p. 224]. But we may show that $C \subset \bar{S}$, where $S$ is the set of $\psi$-singular points of $M$ and that the image of $C$ in $M'$ is therefore a single point. For suppose there exists $x \in C$, $x \notin \bar{S}$. Then there exists an $\epsilon$ such that $\bar{S} \cdot S(x, \epsilon) = 0$. By definition of $M'$, $S(x, \epsilon)$ and $S'(x, \epsilon)$ are topologically equivalent and consequently $S(x', \epsilon)$ contains no $\psi$-singular points, contradicting the fact that $x' \in \bar{S}'$.

* Received June 2, 1941; Presented to the American Mathematical Society September 7, 1937.

[1] A special case of this theorem, namely where $\Gamma$ is the set of all compact metric spaces, has been stated by Whyburn [9].

[2] The argument given by Whyburn for his theorem [9] may be used from here on.

Applications of Theorem 1, for the case where $\Gamma$ is the class of all compact metric continua, have already been pointed out by Whyburn [9]. In particular (for the present paper the most interesting case), if $\psi$ is the property of local connectedness $(0-lc)$ and $\Gamma$ is the class of all compact metric continua, then we get the well-known theorem of Moore [7] that the space of "prime parts" of a compact metric continuum is peanian. An elementary application to the class $\Gamma$ of all compact metric spaces is that where $\psi$ denotes 0-dimensionality; we then get the theorem of Alexandroff [1] that the set of components of a compact metric space is itself a 0-dimensional space (whose metric may be taken as the distance between components).

Now it is to be noted that the $0-lc$ property is not expansive relative to the class of all compact metric spaces (as for instance in the space constituted by the cartesian points $(0,0)$ and $(1/n, 0)$, $n = 1, 2, 3, \cdots$), but is expansive relative to the class of compact metric spaces whose 0-dimensional Betti numbers are finite. By making the convention that an $lc^{-1}$ is simply a compact metric space with no local connectedness[3] properties assigned, the fact just stated comes under the case $n = -1$ of the following theorem:

THEOREM 2. *The property $\psi_{n+1}$ of being $(n+1)-lc$ is expansive relative to the class $C_n^{n+1}$ of compact metric spaces that are $lc^n$ and have finite $(n+1)$-dimensional Betti numbers. Moreover, every $\psi_{n+1}$-singular point of an element $M$ of $C_n^{n+1}$ is in a non-degenerate component of the closure of the set of $\psi_{n+1}$-singular points of $M$.*

*Proof.* Let $M \epsilon C_n^{n+1}$ and $p \epsilon M$ be a point at which $M$ is not $(n+1)-lc$. By Lemma $3^{n+1}$ of [11], there exists an $\epsilon > 0$ such that if $U$, $V$, $W$ are neighborhoods of $p$ satisfying the relations $S(p,\epsilon) \supset U \supset \bar{V}$, $V \supset \bar{W}$, then on the boundary of $V$ there are infinitely many $(n+1)$-cycles independent relative to homologies in $\bar{U} - W$.

Denoting by $S$ the set of $\psi_{n+1}$-singular points of $M$, suppose that the component of $\bar{S}$ determined by $p$ is $p$ itself. Then there exists a neighborhood $V$ of $p$ such that $S(p,\epsilon) \supset \bar{V}$ and the boundary, $F$, of $V$ contains no points of $\bar{S}$. We may cover $F$ by a finite set of spherical neighborhoods whose closures lie in $S(p,\epsilon)$ and do not meet $\bar{S}$. Denote the open set formed by the sum of these neighborhoods by $O$, and let $U = V + O$, $W = V - V \cdot \bar{O}$. It follows that $F$ is a compact subset of the locally compact metric space $U - \bar{W}$, the

[3] Throughout we use Vietoris cycles usually with a finite or compact topological coefficient group, and local connectedness in the homology sense: For example, a compact set is locally $n$-connected $(n-lc)$ if given $\epsilon > 0$ there exists $\delta > 0$ such that a cycle of diameter $< \delta$ bounds on a set of diameter $< \epsilon$. A set is called $lc^n$ if it is $i - lc$ for $i = 0, 1, \cdots, n$. These properties are localized in the usual way.

latter being $lc^{n+1}$. Consequently by Theorem 2 of [11], at most a finite number of $(n+1)$-cycles of $F$ are independent relative to homologies in $U - \bar{W}$. But this contradicts the definition of $\epsilon$ above.

With the same meanings for $C_n^{n+1}$ and $\psi_{n+1}$ as in Theorem 2, we may state the following theorem:

THEOREM 3. *If $M \epsilon C_n^{n+1}$ and the $\psi_{n+1}$-prime parts of $M$ are simply $i$-connected*[4] *for $i = 1, 2, \cdots, n$, then the space $M'$ of $\psi_{n+1}$-prime parts of $M$ is $lc^{n+1}$.*

*Proof.* By Theorem 2, $\psi_{n+1}$ is expansive relative to the class $C_n^{n+1}$. We shall show that $M' \epsilon C_n^{n+1}$ whereupon the conclusion of the theorem follows from Theorems 1 and 2.

That $p^{n+1}(M')$ is finite follows from a theorem of Vietoris [8] and the fact that $p^{n+1}(M)$ is finite.

Consider $p' \epsilon M'$ and $\epsilon > 0$. Denote the mapping of $M$ into $M'$ defined in the introduction by $f$. Then $U = f^{-1}[S(p', \epsilon)]$ is an open subset of $M$ containing the set $P = f^{-1}(p')$. Let $0 < \delta < \epsilon$ and $V = f^{-1}[\bar{S}(p', \delta)]$; obviously $V$ is a closed subset of $M$ such that $U \supset V$.

In the mapping $\bar{S}(p', \delta) = f(V)$, the counter images of points are all simply $i$-connected for $i = 0, 1, \cdots, n$. Consequently [8] to every cycle $\gamma^i$ of $\bar{S}(p', \delta)$ there corresponds in $\bar{S}(p', \delta)$ a cycle $z^i$ such that $\gamma^i \sim z^i$ in $\bar{S}(p', \delta)$ and $z^i$ is the mapping of a cycle $\Gamma^i$ of $V$. By Theorem 2 of [11] only a finite number of the $i$-cycles of $V$ are independent relative to homologies in $U$, and as the mapping $f$ extends to $f(U) = S(p, \epsilon)$, it follows that only a finite number of cycles of $\bar{S}(p', \delta)$ are independent relative to homologies in $S(p', \epsilon)$ and hence $M'$ is $i - lc$ at $p'$.

COROLLARY. *If $M$ is a Peano space such that $p^1(M)$ is finite, then the space of $\psi_1$-prime parts of $M$ is an $lc^1$.*

Now as a decomposition theorem, Theorem 3 is of somewhat restricted application when $n$ is large. Suppose, however, we consider the property $\psi_0^n$ of being $lc^n$, so that a $\psi_0^n$-singular point is one at which the space fails to be $i - lc$ for some $i \leqq n$.

THEOREM 4. *The property $\psi_0^n$ of being $lc^n$ $(0 \leqq n)$ is expansive relative to the class $C_{-1}^n$ of compact metric spaces that have finite $i$-dimensional Betti numbers for all $i \leqq n$.*

*Proof.* Let $M \epsilon C_{-1}^n$. If $M$ fails to be $0 - lc$, then the closure of the

---

[4] i. e., all the $i$-cycles of each prime part bound thereon.

set of points at which $M$ is not $0 - lc$ contains a continuum; and *a fortiori* the set of $\psi_0{}^n$-singular points contains a continuum. In general, if $M$ fails to have property $\psi_0{}^n$, let $k$ be the largest integer for which $M$ is $lc^k$. Since $k < n$, $p^{k+1}(M)$ is finite by hypothesis and $M \in C_k{}^{k+1}$. Hence by Theorem 2 the closure of the set of points at which $M$ is not $(k+1) - lc$ contains a continuum and *a fortiori* the same is true of the set of $\psi_0{}^n$-singular points of $M$.

THEOREM 5. *If $M \in C_{-1}{}^n$ and the $\psi_0{}^n$-prime parts of $M$ are simply $i$-connected for $i = 1, 2, \cdots, n-1$, then the space $M'$ of $\psi_0{}^n$-prime parts of $M$ is an $lc^n$.*

*Proof.* Since by Theorem 4 $\psi_0{}^n$ is expansive relative to $C_{-1}{}^n$, we need, in order to apply Theorem 1, only show that under the given conditions the space $M' \in C_{-1}{}^n$. The latter follows immediately from results due to Vietoris [8].

COROLLARY. *If $M \in C_{-1}{}^1$, then the space of $\psi_0{}^1$-prime parts of $M$ is an $lc^1$.*

We are now in a position to state our principal theorem. Hereafter if $r$ and $n$ are integers, we denote by $C_r{}^n$ the class of all compact metric spaces $M$ such that $M$ is $lc^r$ and the Betti numbers $p^i(M)$ are finite for $r < i \leqq n$. And by $\psi_r{}^n$ we denote the property of being $i - lc$ for $i = r, r+1, \cdots, n$.

PRINCIPAL THEOREM. *Let $r$ and $n$ be integers such that $-2 < r < n$. Then the property $\psi_{r+1}{}^n$ is expansive relative to the class of spaces $C_r{}^n$. If $M \in C_r{}^n$ and the $\psi_{r+1}{}^n$-prime parts of $M$ are simply $i$-connected for $i = 1, 2, \cdots, n-1$, then the space of $\psi_{r+1}{}^n$-prime parts of $M$ is an $lc^n$.*

That $\psi_{r+1}{}^n$ is expansive relative to $C_r{}^n$ is proved as in Theorem 4, starting inductively with the result of Theorem 4 $(r = -1)$. To prove the remainder of the theorem we may apply the results of Vietoris on continuous mappings [8] as in the proofs of Theorems 3 and 5 to show that under the given conditions the space $M'$ of $\psi_{r+1}{}^n$-prime parts of $M$ is an element of $C_r{}^n$; then $M'$ is an $lc^n$ by Theorem 1.

## SOME APPLICATIONS

In the article of Hahn [3] in which what we call the $\psi_0{}^0$-prime parts were introduced, it was shown that a continuum which is irreducible between two of its points and has more than one prime part is a simple arc of prime parts. In R. L. Moore's investigations, besides the results already cited above, some attention was paid to the case where the space of prime parts is a simple closed curve [6]. It would be of interest to determine under what conditions the $\psi_r{}^n$-prime parts of a space constitute spaces falling among the more well-known topological configurations of higher dimensions. The following two theorems illustrate this type of application.

THEOREM 6. *Let $M$ be a space such that $p^2(M) > 0$ irreducibly, $p^1(M)$ is finite, and the $\psi_0{}^1$-prime parts of $M$ are simply* 1*-connected and none forms a* 1*-barrier.*[5] *Then the space of $\psi_0{}^1$-prime parts of $M$ is either a closed* 2*-dimensional manifold or a point.*

*Proof.* I have shown in unpublished work that if a compact metric space $M'$ has $p^2(M') > 0$ irreducibly, and is both semi-1-connected [11, p. 549] and locally 1-avoidable [12], then $M'$ is a closed 2-dimensional manifold. We shall show that the space $M'$ of $\psi_0{}^1$-prime parts of $M$, if not a single point, has these properties.

By the Corollary of Theorem 5, $M'$ is an $lc^1$, hence *a fortiori* semi-1-connected. As the $\psi_0{}^1$-prime parts are simply 1-connected by hypothesis, and simply 2-connected when $M$ has more than one such prime part, it follows [8] that $p^2(M') = p^2(M) > 0$. Suppose $M'$ has a proper subset $F'$ such that $p^2(F') > 0$. Let $\gamma^2$ be a non-bounding cycle of $F'$. Then [8] there exists on $F'$ a cycle $z^2$ such that $\gamma^2 \sim z^2$ on $F'$ and $f^{-1}(z^2)$ is a cycle of $F = f^{-1}(F')$ (where $f$ is the mapping of $M$ into $M'$). But $F$ is a proper subset of $M$ and consequently $f^{-1}(z^2) \sim 0$ on $F$. But then $z^2 \sim 0$ on $F'$ and $\gamma^2 \sim 0$ on $F'$. Thus $p^2(M') > 0$ irreducibly.

Consider any $p' \epsilon M'$, $\epsilon > 0$, and let $0 < \eta < \delta < \epsilon$ such that 1-cycles of $F(p', \delta)$ bound in $S(p', \epsilon)$. As $M'$ is $lc^1$ at most a finite number of 1-cycles of $F(p', \delta)$ fail to bound in $S(p', \epsilon) - \bar{S}(p', \eta)$. Let $\gamma^1$ be such a cycle. Let $p = f^{-1}(p')$ and $F = f^{-1}[F(p', \delta)]$. There exists on $F(p', \delta)$ a cycle $z^1$ such that $\gamma^1 \sim z^1$ on $F(p', \delta)$ and $\Gamma^1 = f^{-1}(z^1)$ is a cycle of $F$ [8]. As $\gamma^1 \sim 0$ in $S(p', \epsilon)$, $z^1 \sim 0$ in $S(p', \epsilon)$, hence $\Gamma^1 \sim 0$ on $M$ [8]. As $p$ is not a barrier of $\Gamma^1$, then $\Gamma^1 \sim 0$ on some closed subset $K$ of $M - p$. But then $z^1 \sim 0$ on $f(K) \subset M' - p'$, implying $\gamma^1 \sim 0$ on $M' - p'$. It follows that $\eta$ can be so chosen that for every $\gamma^1$ of $F(p', \delta)$, $\gamma^1 \sim 0$ on $M' - S(p', \eta)$ and $M'$ is locally 1-avoidable. This completes the proof.

For the case $n > 2$ we have the following theorem:

THEOREM 7. *Let $M \epsilon C^{n-1}_{-1}$, $p^n(M) = 1$ and $p^n(F) = 0$ if $F$ is a proper subset of $M$. Then if the $\psi_0{}^{n-1}$-prime parts of $M$ are simply $i$-connected for $i = 1, 2, \cdots, n-1$, and none is an $i$-dimensional local barrier for $i = 1, 2, \cdots, n-2$ nor an $(n-1)$-barrier of $M$, then the space $M'$ of such prime parts is either a generalized closed $n$-manifold* [10] *or a point.*

*Proof.* Suppose $M'$ is non-degenerate. Then the $\psi_0{}^{n-1}$-prime parts of $M$

[5] i. e., if $P$ is a $\psi_0{}^1$-prime part and $z^1$ is a 1-cycle which lies on $M - P$ and bounds on $M$, then $z^1$ bounds on $M - P$.

are simply $i$-connected for all $i \leqq n$ and $p^n(M') = 1$ [8]. And if $F'$ is a proper subset of $M'$, $p^n(F') = p^n[f^{-1}(F')] = 0$.

That $M'$ is $lc^{n-1}$ follows from Theorem 5. Hence if $p' \epsilon M'$, $0 < \eta < \delta < \epsilon$, the number of $i$-cycles $(i \leqq n-1)$ of $F(p', \delta)$ independent relative to homologies in $S(p', \epsilon) - \bar{S}(p', \eta)$ is finite [11]. The local $(n-1)$-avoidability of $M'$ now follows as in the proof of the local 1-avoidability in the preceding theorem.

Consider the case $1 \leqq i \leqq n-2$. For any $\epsilon > 0$ the set $U = f^{-1}[S(p', \epsilon)]$ is an open subset of $M$ containing $P = f^{-1}(p')$. By hypothesis there exists a $\delta_1 > 0$ such that every $i$-cycle of $\bar{S}(P, \delta_1) - P$ bounds on $U - P$, where $\bar{S}(P, \delta_1) \subset U$. By continuity considerations there exists a $\delta'$ such that $V = f^{-1}[\bar{S}(p', \delta')] \subset S(P, \delta_1)$. Consider any cycle $\gamma^i$ of $F(p', \delta')$. It is homologous on $F(p', \delta')$ to a cycle $z^i$ such that $f^{-1}(z^i)$ is a cycle $\Gamma^i$ of $f^{-1}[F(p', \delta')]$. As $\Gamma^i$ bounds on $U - P$, $z^i$ bounds on $S(p', \epsilon) - p'$ and hence $\gamma^i$ bounds on $S(p', \epsilon) - p'$. From the observation stated in the second sentence of the preceding paragraph it now follows that there exists an $\eta'$ such that $0 < \eta' < \delta'$ and every $i$-cycle of $F(p', \delta')$ bounds on $S(p', \epsilon) - \bar{S}(p', \eta')$. Thus $M$ is a generalized closed $n$-manifold.

These theorems have applications in turn to subsets of the euclidean $n$-sphere $S^n$. For example, a set $M$ satisfying the hypotheses of Theorem 6 and lying in $S^3$ will, in case its $\psi_0{}^1$-singular points are not dense in it, have exactly two complementary domains and be their common boundary. (In this connection see [5, 6]).

Evidently a theorem similar to Theorem 7 is derivable from the Principal Theorem; i. e., *Theorem 7 remains true if we replace* $C^{n-1}_{-1}$ *by* $C_r{}^n$ *and* $\psi_0{}^{n-1}$ *by* $\psi_{r+1}{}^n$.

In closing we remark that in a paper published in 1930 [2] Aronszajn has determined conditions on the class of sets which serve to define a given local topological property $\psi$, that are sufficient to ensure that every point of the set $S$ of $\psi$-singular points of a compact metric continuum should lie in a non-degenerate continuum of $S$. Since application of Theorem 1 above requires only the existence of *at least one* non-degenerate component *in the closure* of $S$, it might be of interest to investigate general conditions sufficient to ensure expansiveness as we have defined it. It is clear from the above results that in any such investigation the problem is one of adjusting the conditions on the class of sets defining the topological property to the class of spaces under consideration (the latter were continua in the Aronszajn investigation).

UNIVERSITY OF MICHIGAN.

## BIBLIOGRAPHY

1. P. Alexandroff, "Über stetige Abbildungen kompakter Räume," *Koninklijke Akademie van Wetenschappen te Amsterdam*, Proceedings of the Section of Sciences, vol. 28 (1925), pp. 997-999.

2. N. Aronszajn, "Einige Bemerkungen über den Begriff des lokalen Zusammenhanges," *Monatshefte für Mathematik und Physik*, vol. 37 (1930), pp. 241-252.

3. H. Hahn, "Über irreduzible Kontinua," *Sitzungsberichte der Akademie der Wissenschaften in Wien*, Mathematisch-Naturwissenschaftlichen Klasse, Abt. IIa, vol. 130 (1921), pp. 217-250.

4. F. Hausdorff, "*Mengenlehre*," Leipzig, 1935, pp. 223-225.

5. R. L. Moore, "Concerning the common boundary of two domains," *Fundamenta Mathematicae*, vol. 6 (1924), pp. 203-213.

6. ———, "Concerning the prime parts of certain continua which separate the plane," *Proceedings of the National Academy of Sciences*, vol. 10 (1924), pp. 170-175.

7. ———, "Concerning the prime parts of a continuum," *Mathematische Zeitschrift*, vol. 22 (1925), pp. 307-315.

8. L. Vietoris, "Über den höheren Zusammenhang kompakter Räume und eine Klasse von zusammenhangstreuen Abbildungen," *Mathematische Annalen*, vol. 97 (1927), pp. 454-472.

9. G. T. Whyburn, "A decomposition theorem for closed sets," *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 95-96.

10. R. L. Wilder, "Generalized closed manifolds in $n$-space," *Annals of Mathematics*, (2), vol. 35 (1934), pp. 876-903.

11. ———, "On locally connected spaces," *Duke Mathematical Journal*, vol. 1 (1935), pp. 543-555.

12. ———, "Sets which satisfy certain avoidability conditions," *Časopis pro Pěstovani Matematiky a Fysiky* (1938), pp. 185-198.

# ON FOURIER AVERAGES.*

By Aurel Wintner.

---

If $g(t), -\infty < t < \infty$, is integrable on every finite $t$-interval, let $M_t\{g(t)\}$ denote the average

$$M_t\{g(t)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(t)\,dt,$$

provided that this limit exists.

If $g(t) \geqq 0$, put

$$\bar{M}_t\{g(t)\} = \overline{\lim_{T\to\infty}} \frac{1}{2T} \int_{-T}^{T} g(t)\,dt$$

(even if $\bar{M}_t\{g(t)\} = \infty$).

If, for a given $f(t), -\infty < t < \infty$, the average

$$M_t\{f(t+s)\bar{f}(t)\}$$

exists for $-\infty < s < \infty$ and represents a continuous function of $s$, then $f(t)$ is said to possess an auto-correlation.

For a fixed real $\lambda$, the $\lambda$-th Fourier average of an $f(t), -\infty < t < \infty$, is defined by

$$c(\lambda) = M_t\{e^{-i\lambda t} f(t)\},$$

provided that this average exists.

It was proved by Karamata [1] that, if $\bar{M}_t\{|f(t)|^2\} < \infty$ and if $c(\lambda)$ exists for *every* $\lambda$, then the function $|c(\lambda)|^2, -\infty < \lambda < \infty$, is of bounded variation and vanishes nearly everywhere (that is, except for a set of $\lambda$-values which is enumerable, at most). Karamata's proof of his theorem is quite elaborate. The following, more general, theorem is an immediate consequence of an adaptation of the classical argument of Bessel:

(I) *If the upper mean square, $\bar{M}_t\{|f(t)|^2\}$, of a measurable function $f(t), -\infty < t < \infty$, is finite, then the set of those values $\lambda$ for which the Fourier average, $c(\lambda)$, of $f(t)$ exists but does not vanish is enumerable, at most; furthermore, if $\lambda_1, \cdots, \lambda_n$ is any set of $\lambda$-values for which $c(\lambda)$ exists, then*

---

* Received February 26, 1941.

[1] J. Karamata, "Sur la moyenne arithmétique des coéfficients d'une série de Taylor," *Mathematica* (Cluj), vol. 1 (1929), pp. 99-106.

$$\bar{M}_t\{|f(t)|^2\} - \sum_{m=1}^{n} |c(\lambda_m)|^2 \geqq 0. \tag{1}$$

In fact, since

$$\bar{M}_t\{|f(t) - \sum_{m=1}^{n} c(\lambda_m)e^{i\lambda_m t}|^2\} \tag{2}$$

is non-negative, it is sufficient to prove that the expression (2) is identical with the difference (1), if $c(\lambda)$ exists for $\lambda = \lambda_1, \cdots, \lambda_n$. Furthermore, it may be assumed in the proof of this identity that $f(t)$ is real-valued. Then, if $c(\lambda)$ exists, $c(-\lambda)$ exists also, it being the complex conjugate of the number $c(\lambda)$.

Since $\bar{M}_t\{|f(t)|^2\} < \infty$ is a constant, and since $c(\pm\lambda_m)$ exists for $m = 1, \cdots, n$, the absolute square beneath the $\bar{M}_t\{\cdots\}$-sign of (2) is the sum of $(n+1)^2$ terms, $(n+1)^2 - 1$ of which possess averages $M_t\{\cdots\}$ (and not merely finite upper averages $\bar{M}_t\{\cdots\}$); while the $(n+1)^2$-th term has the upper average $\bar{M}_t\{|f(t)|^2\}$. Since this implies that the expression (2) is identical with the difference (1), the proof of (I) is complete.

In view of (I), it is natural to ask whether an unspecified measurable $f(t)$ can or cannot possess a non-enumerable set of non-vanishing Fourier averages $c(\lambda)$, if either the assumption, $\bar{M}_t\{|f(t)|^2\} < \infty$, of (I) is completely omitted or is replaced by the weaker assumption $\bar{M}_t\{|f(t)|\} < \infty$. The answer is not known to be affirmative; not even under the additional assumption that *all* the Fourier averages $c(\lambda)$ of $f(t)$ exist.[2]

The following considerations deal with the question of existence of the Fourier averages. More specifically, a sufficient condition will be obtained for those functions $f(t)$ for which the Fourier average $c(\lambda)$ exists for every $\lambda$ not contained in a $\lambda$-set of measure zero. The method to be applied consists of an adaptation of a procedure by means of which certain considerations of Riemann were legalized recently.[3] The excluded zero set will not, in general, be enumerable; in other words, $c(\lambda)$ will exist "almost everywhere" but not necessarily "nearly everywhere." This circumstance makes the result to be obtained irrelevant for the purposes of an harmonic analysis of ergodic systems.[4] In fact, for these purposes the following theorem was needed:

(II) *If $f(t)$ possess an auto-correlation, then the Fourier average $c(\lambda)$*

---

[2] Cf. A. Wintner, "On the distribution function of the remainder term of the prime number theorem," *American Journal of Mathematics*, vol. 63 (1941), pp. 233-248.

[3] A. Wintner, "On Riemann's fragment concerning elliptic modular functions," *American Journal of Mathematics*, vol. 63 (1941), pp. 628-634.

[4] N. Wiener and A. Wintner, "Harmonic analysis and ergodic theory," *American Journal of Mathematics*, vol. 63 (1941), pp. 415-426.

*of $f(t)$ exists for nearly all $\lambda$, provided that $f(t)$ is either a bounded function or satisfies certain, more general, Tauberian conditions.*

It is not known whether or not the assertion of (II) remains true if no Tauberian proviso is made. No such proviso is made by the following lemma, which assumes very much less than (II) but makes a weaker statement than (II):

(III) *If $\bar{M}_t\{|f(t)|^2\} < \infty$, then $c(\lambda)$ exists for almost all $\lambda$.*

(III) was proved by Karamata[1] for the case of a sequence, where $f(t) = f(n)$ for $n < t \leqq n+1$; $n = 0, \pm 1, \pm 2, \cdots$. He first verified the corresponding Abelian fact and then applied the fundamental Tauberian theorem of Hardy and Littlewood concerning the equivalence of the summation processes of Abel and Cesàro in the slowly oscillating case. However, it turns out that by a simpler procedure much more than (III) can be established:

(IV) *If $f(t)$ is integrable on every finite $t$-interval, and if there exists a sufficiently small $\epsilon > 0$ satisfying*

$$\int_{-T}^{T} |f(t)|^2 dt = O(T^{2-\epsilon}) \text{ as } T \to \infty, \tag{3}$$

*then the Fourier average, $c(\lambda)$, of $f(t)$ exists and vanishes for almost all $\lambda$.*

Notice that (I) is applicable to both (II) and (III) but not to (IV); in fact, the assumption of (III) is the particular case $\epsilon = 1$ of the assumption (3) of (IV).

If $h(t)$, $0 \leqq t < \infty$, is a function for which

$$\int_0^T th(t)\,dt$$

exists for every $T > 0$ and tends, as $T \to \infty$, to a finite limit, then, since the mean-value

$$\frac{1}{T}\int_0^T t\Big(\int_0^t sh(s)\,ds\Big)dt$$

tends to the same limit, a partial integration shows that

$$\frac{1}{T}\int_0^T h(t)\,dt \to 0 \text{ as } T \to \infty.$$

Thus it is clear that, if $g(t)$, $-\infty < t < \infty$, is a function which is integrable on every finite $t$-interval and has the property that

$$\Big(\int_{-T}^{T} - \int_{-1}^{1}\Big)\frac{g(t)}{t}\,dt$$

tends to a finite limit as $T \to \infty$, then

$$\frac{1}{2T}\int_{-T}^{T} g(t)\,dt \to 0 \text{ as } T \to \infty,$$

which means that $M_t\{g(t)\}$ exists and vanishes.

This Abelian remark, when applied to the function $g(t) = e^{-i\lambda t}f(t)$, where $\lambda$ is fixed, implies that (IV) is a corollary of the following theorem:

(V) *If a function $f(t)$, $-\infty < t < \infty$, is integrable on every finite $t$-interval, and if there exists a sufficiently small $\epsilon > 0$ satisfying* (3), *then the improper integral*

$$\left(\int_{-\infty}^{\infty} - \int_{-1}^{1}\right)\frac{f(t)}{t}\,e^{-i\lambda t}dt \equiv \lim_{T\to\infty}\left(\int_{-T}^{T} - \int_{-1}^{1}\right)\frac{f(t)}{t}\,e^{-i\lambda t}dt$$

*is convergent for almost all $\lambda$.*

The proof of (V) can be based on the following theorem of Plancherel:[5] If a function $k(t)$ is not only of class $(L^2)$ on the interval $-\infty < t < \infty$ but satisfies the additional restriction

$$\left(\int_{-\infty}^{\infty} - \int_{-1}^{1}\right) | k(t) \log | t | |^2 dt < \infty,$$

then the Fourier transform

$$\int_{-\infty}^{\infty} e^{iut}k(t)\,dt$$

is not only convergent in the mean $(L^2)$ but exists for almost all $u$.

On placing in this theorem

$$-u = \lambda \text{ and } k(t) = f(t)/t \qquad (\text{if } |t| > 1),$$

we see that, in order to prove (V), it is sufficient to show that

$$\left(\int_{-\infty}^{\infty} - \int_{-1}^{1}\right)\left|\frac{f(t)}{t}\log|t|\right|^2 dt < \infty$$

by virtue of (3). In other words, it is sufficient to show that, if (3) is satisfied,

$$\int_{1}^{T} |f(t)|^2 \frac{\log^2 t}{t^2}\,dt = O(1) \text{ as } T \to \infty.$$

Hence, partial integration shows that it is sufficient to ascertain that

---

[5] M. Plancherel, "Sur la convergence et la sommation par les moyennes de Cesàro . . .," *Mathematischen Annalen*, vol. 76 (1915), pp. 315-326.

$$O(T^{2-\epsilon})\,\frac{\log^2 T}{T^2} - \int_1^T O(t^{2-\epsilon})\,\frac{d}{dt}\,\frac{\log^2 t}{t^2}\,dt = O(1) \text{ as } T \to \infty.$$

But this is obvious from

$$\frac{d}{dt}\,\frac{\log^2 t}{t^2} = 2\,\frac{\log t}{t}\cdot\frac{1-\log t}{t^2} = O\left(\frac{\log^2 t}{t^3}\right) = O\left(\frac{t^{\frac{1}{2}\epsilon}}{t^3}\right).$$

This completes the proof of (V).

An Abelian corollary of the Abelian corollary, (IV), of (V) may be formulated as follows:

(VI) *If a sequence of constants $a_1, a_2, \cdots$ satisfies the estimate*

$$\sum_{m=1}^{n} |\,a_m\,|^2 = O(n^{2-\epsilon}), \qquad n \to \infty, \tag{4}$$

*for a sufficiently small $\epsilon > 0$, then the regular function defined by the power series*

$$F(z) = \sum_{n=1}^{\infty} a_n z^n, \qquad |\,z\,| < 1, \tag{5}$$

*behaves in the neighborhood of the circle $|\,z\,| = 1$ in such a way that*

$$F(re^{i\phi}) = o(1-r)^{-1}, \qquad r \to 1, \tag{6}$$

*holds for almost all $\phi$; furthermore, this estimate remains valid for almost all $\phi$, if the radial approach is replaced by Stolz paths.*

In fact, if $c_1, c_2, \cdots$ is a sequence of numbers for which the arithmetical mean tends to a finite limit

$$\lim_{n\to\infty} \frac{c_1 + \cdots + c_n}{n} = C,$$

then, according to the Abelian theorem of Frobenius,

$$\lim_{r\to 1}\,(1-r)\sum_{n=1}^{\infty} c_n r^n = C, \qquad (r < 1);$$

furthermore,

$$\lim_{x\to 1}\,(1-x)\sum_{n=1}^{\infty} c_n x^n = C,$$

if $x$ is complex but is restricted to Stolz paths leading to $x = 1$. Hence, on choosing

$$C = 0 \text{ and } c_n = a_n e^{in\phi},$$

where $\phi$ is fixed, we see that, in order to prove (VI), it is sufficient to verify the following statement: If a sequence $a_1, a_2, \cdots$ satisfies (4) for a sufficiently small $\epsilon > 0$, then

$$\lim_{n\to\infty} \frac{a_1 e^{i\phi} + \cdots + a_n e^{in\phi}}{n} = 0$$

holds for almost all $\phi$. But this statement is equivalent to the one which results on formulating (IV) for sequences, instead of for functions.

As an application of (VI), consider Lambert's own series,[6]

$$(8) \qquad L(z) = \sum_{n=1}^{\infty} \frac{z^n}{1 - z^n} \equiv \sum_{n=1}^{\infty} \tau(n) z^n, \qquad (|z| < 1),$$

where $\tau(n)$ denotes the number of the divisors of $n$. Thus

$$(9) \qquad 1 \leqq \tau(n) \neq O(1)$$

and, as is well-known,

$$(10) \qquad \tau(n) = O(n^{\delta}) \text{ for every fixed } \delta > 0.$$

It is also known that the circle $|z| = 1$ is a natural boundary[7] of (8), and that the increase of (8) in the steepest direction, $\arg z = 0$, is characterized by[8]

$$L(r) \sim \frac{1}{1-r} \log \frac{1}{1-r}, \qquad r \to 1 - 0.$$

However, the increase of (8) in the direction of a "general" $\phi = \arg z$ is less steep; in fact one has, *for almost all* $\phi$,

$$L(re^{i\phi}) = O\left(\frac{1}{1-r}\right) \text{ and even } L(re^{i\phi}) = o\left(\frac{1}{1-r}\right)$$

(if not a still better estimate).

This is implied by (VI), since, as is seen from (10), condition (4) is satisfied for $a_n = \tau(n)$ by certain $\epsilon > 0$ (in fact, by every positive $\epsilon < 1$). On the other hand, (9) shows that Karamata's condition, that is, the case $\epsilon = 1$ of (4), is not satisfied.

For more general Lambert series, (VI) leads to the following result:

(VII) *If a sequence of numbers $c_1, c_2, \cdots$ satisfies*

$$(11) \qquad c_n = O(n^{\frac{1}{2}-\eta}) \textit{ for a sufficiently small } \eta > 0,$$

*then the boundary behavior of the regular function which is defined for $|z| < 1$ by the Lambert series*

---

[6] J. H. Lambert, *Anlage zur Architektonik* . . ., vol. 2 (1771), § 875.

[7] This is implied by Carlson's theorem on power series with integral coefficients but can be verified directly.

[8] Cf. the end of the paper by É. le Roy, "Valeurs asymptotiques des certaines séries . . .," *Bulletin des Sciences Mathématiques*, ser. 2, vol. 24 (1900), pp. 245-268.

$$F(z)=\sum_{n=1}^{\infty} c_n \frac{z^n}{1-z^n} \tag{12}$$

*is subject to* (6) *for almost all* $\phi$; *furthermore, the same estimate holds, for almost all* $\phi$, *even if the radial approach is replaced by Stolz paths.*

In fact, (11) insures that (12) represents a regular function for $|z|<1$. Furthermore, it is clear that the coefficients, $a_n$, of the power series, (5), of (12) are given by

$$a_n=\sum_{d|n} c_d,$$

where $d$ runs through all the divisors of $n$; so that, from (11),

$$a_n=O(n^{\frac{1}{2}-\eta})\sum_{d|n} 1\equiv O(n^{\frac{1}{2}-\eta})\tau(n)=O(n^{\frac{1}{2}-\eta+\delta}).$$

Since $\eta>0$ and $\delta>0$ can be chosen arbitrarily, it follows, by choosing $0<\frac{1}{2}-\eta+\delta<\frac{1}{2}$, that (4) is satisfied by some $\epsilon>0$. Hence, (VII) is implied by (VI).

The only result known in the direction of (VII) seems to be an Abelian theorem of Knopp and Landau,[9] which states that, if the coefficients of the Lambert series (12) satisfy the drastic assumption

$$\sum_{n=1}^{\infty}\frac{|c_n|}{n}<\infty, \tag{13}$$

then

$$\lim_{r\to 1}(1-r)\sum_{n=1}^{\infty} c_n\frac{(re^{i\phi})^n}{1-(re^{i\phi})^n}=\sum_{n=1}^{\infty}\lim_{r\to 1}(1-r)c_n\frac{(re^{i\phi})^n}{1-(re^{i\phi})^n},$$

along with the corresponding relation for Stolz paths, holds for every angle $\phi$; so that (6) holds for every irrational $\phi/\pi$.

It is interesting that, roughly speaking, (13) is more stringent than (11) by the order of $\sqrt{n}$. In particular, (13) is violated even by Lambert's own series (8), where $a_n=1$. However, it does not appear to be known whether or not (6) is true for all irrational $\phi/\pi$, if (12) is given by (8). In fact, the problem of the "heaviest singularities," being an arithmetical issue of "major and minor arcs," requires to-day an explicit analytical technique of the "nearly all," and not the above Lebesgue approach, which leads only to the unspecified "almost all." Correspondingly, the order $\sqrt{n}$ mentioned above can be thought of as a manifestation of the standard improvement of the apparent order in case of random distributions.

THE JOHNS HOPKINS UNIVERSITY.

---

[9] For a proof much simpler than that of these authors, cf. H. Späth, "Über Lambertsche Reihen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 481-486.

# ON METHODS OF SUMMABILITY AND MASS FUNCTIONS DETERMINED BY HYPERGEOMETRIC COEFFICIENTS.*

By Ralph Palmer Agnew.

**1. Introduction.** We begin by introducing pertinent notation and results from a theory originated by Hurwitz and Silverman [1] and by Hausdorff.[2] A sequence-to-sequence transformation of the form

$$A: \qquad \sigma_n = \sum_{k=0}^{\infty} a_{nk} s_k$$

commutes with (or is permutable with) the arithmetic mean transformation $C_1$ if and only if it has the form

$$A(\lambda): \qquad \sigma_n = \sum_{k=0}^{n} \binom{n}{k} \left\{ \sum_{j=k}^{n} (-1)^{j-k} \binom{n-k}{j-k} \lambda_j \right\} s_k$$

where $\lambda_0, \lambda_1, \cdots$ is a sequence of complex constants. The sum in braces is, in difference notation, $\Delta^{n-k}\lambda_n$. If $A'$ and $A''$ are two transformations of the form $A(\lambda)$ generated respectively by two sequences $\lambda_n'$ and $\lambda_n''$, then the product transformation $A^* = A'A''$ is generated by the sequence $\lambda_n^* = \lambda_n'\lambda_n''$; and the extension to products of three or more factors is obvious. If $\lambda_n \neq 0$ for each $n$, then $A(\lambda)$ has an inverse generated by the sequence $\lambda_n^{-1}$ of reciprocals of the elements of the sequence $\lambda_n$. One of the main reasons why transformations which commute with $C_1$ are of importance lies in the fact that two transformations which commute with $C_1$ commute with each other; this fact is proved and its significance is emphasized both by Hurwitz and Silverman and by Hausdorff.

The transformation $A(\lambda)$ is conservative [3] if and only if a mass function $\chi(t)$ exists such that

$$(1) \qquad \chi(0) = 0$$

$$(2) \qquad \chi(t) \text{ has bounded variation over } 0 \leqq t \leqq 1$$

* Received December 30, 1940; Presented to the American Mathematical Society December 30, 1940.

[1] W. A. Hurwitz and L. L. Silverman, "On the consistency and equivalence of certain definitions of summability," *Transactions of the American Mathematical Society*, vol. 18 (1917), pp. 1-20.

[2] F. Hausdorff, "Summationsmethoden und Momentfolgen, I and II," *Mathematische Zeitschrift*, vol. 9 (1921), pp. 74-109 and 280-299.

[3] That is, such that the existence of $\lim s_k$ implies the existence of $\lim \sigma_n$.

$$(3) \qquad \lambda_n = \int_0^1 t^n d\chi(t) \qquad (n = 0, 1, 2, \cdots);$$

and is regular[4] if and only if $\chi(t)$ exists such that (1), (2), and (3) hold and also

$$(4) \qquad \chi(1) = 1$$

$$(5) \qquad \chi(t) \text{ is continuous at } t = 0.$$

A mass function $\chi(t)$ satisfying (1), (2), (4), and (5) is called a *regular mass function*; and a sequence $\lambda_n$ for which a regular mass function exists such that (3) holds is called a *regular moment sequence.*

The Cesàro method $C_r$, $r$ being a complex constant different from $-1, -2, \cdots$, is generated by the sequence

$$(6) \qquad \lambda_n^{(r)} = 1 / \binom{n+r}{r}.$$

The method $C_r$ is regular if and only if $r = 0$ or $r' = Rr > 0$. Here and hereafter $z'$ is used to denote the real part of a complex number $z$. The identity transformation $C_0$ is generated by the sequence $\lambda_n^{(0)} = 1$, $n = 0, 1, 2, \cdots$. The Hölder methods $H_r$ are generated by the sequences $(n+1)^{-r}$ and have the property $H_r H_s = H_{r+s}$. The well-known fact that the Cesàro method $C_r$ and the Hölder method $H_r$ are equivalent when $r' > -1$ may be expressed by writing $C_r \approx H_r$. Since the Cesàro and Hölder methods $C_r$ and $H_s$ have the form $A(\lambda)$ and therefore commute, we can use this equivalence to obtain, when $r', s', r' + s' > -1$,

$$C_r C_s \approx H_r C_s = C_s H_r \approx H_s H_r = H_{r+s} \approx C_{r+s};$$

this implies that

$$(7) \qquad C_r^{-1} C_{r+s} \approx C_s \qquad (r', s', r' + s' > -1).$$

It is a corollary of this result that $C_\alpha \subset C_\beta$ if $-1 < \alpha' < \beta'$.

**2. Hypergeometric summability.** Corresponding to each set of complex constants $\alpha, \beta, \gamma$ for which $\gamma \neq 0, -1, -2, \cdots$ let a sequence $\lambda_n(\alpha, \beta, \gamma)$ be defined by $\lambda_0(\alpha, \beta, \gamma) = 1$ and, when $n = 1, 2, 3, \cdots$,

$$(8) \quad \lambda_n(\alpha, \beta, \gamma) = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)\beta(\beta+1)\cdots(\beta+n-1)}{\gamma(\gamma+1)\cdots(\gamma+n-1)\, 1 \cdot 2 \cdots n}.$$

These $\lambda$'s are the coefficients in the familiar power series expansion $\Sigma \lambda_n(\alpha, \beta, \gamma) z^n$ of the hypergeometric function $F(\alpha, \beta, \gamma; z)$. The sequence

[4] That is, such that the existence of $\lim s_k$ implies $\lim \sigma_n = \lim s_k$.

$\lambda_n(\alpha,\beta,\gamma)$ determines a transformation $A(\lambda_n(\alpha,\beta,\gamma))$ which defines the *hypergeometric method of summability* denoted by $(H,\alpha,\beta,\gamma)$.[5] Using (8) and (6) we find that when $\alpha,\beta,\gamma \neq 0,-1,-2,\cdots$

$$\lambda_n(\alpha,\beta,\gamma) = \binom{n+\alpha-1}{\alpha-1}\binom{n+\beta-1}{\beta-1}\Big/\binom{n+\gamma-1}{\gamma-1}$$
$$= [\lambda_n^{(\alpha-1)}]^{-1}[\lambda_n^{(\beta-1)}]^{-1}[\lambda_n^{(\gamma-1)}]$$

so that use of the theory of § 1 gives the identity

$$(9) \qquad (H,\alpha,\beta,\gamma) = C_{\alpha-1}^{-1}C_{\beta-1}^{-1}C_{\gamma-1} \qquad (\alpha,\beta,\gamma\neq 0,-1,-2,\cdots).$$

The order of the factors on the right is immaterial since the factors commute. So much is known about the Cesàro methods $C_r$, especially when $r' > -1$, that formula (9) and formula (12) which we obtain below furnish an easy method of determining properties of $(H,\alpha,\beta,\gamma)$. In particular the results of the section on hypergeometric summability, pp. 195-201, in the paper of Garabedian and Wall become more comprehensible when interpreted in the light of (9) and (12).

Using (9) and (7), we can show that

$$(10) \qquad C_{\alpha-1}^{-1}C_{\beta-1}^{-1}C_{\gamma-1} = C_{\gamma-\alpha-\beta+1} \qquad (\alpha',\beta',\gamma' > 0;\ \gamma' > \alpha'+\beta'-2);$$

in case $\beta',\gamma' > 0$; $\gamma' > \alpha'+\beta'-2$; and $0 < \alpha' \leqq 1$ this follows from

$$C_{\alpha-1}^{-1}C_{\gamma-1}C_{\beta-1}^{-1} \approx C_{\gamma-\alpha}C_{\beta-1}^{-1} \approx C_{\gamma-\alpha-\beta+1};$$

and in case $\beta',\gamma' > 0$; $\gamma' > \alpha'+\beta'-2$; and $\alpha' > 1$ it follows from

$$C_{\beta-1}^{-1}C_{\gamma-1}C_{\alpha-1}^{-1} \approx C_{\gamma-\beta}C_{\alpha-1}^{-1} \approx C_{\gamma-\alpha-\beta+1}.$$

Combining (9) and (10) we see that if

$$(11) \qquad \alpha',\beta',\gamma' > 0; \qquad \gamma' > \alpha'+\beta'-2,$$

then

$$(12) \qquad (H,\alpha,\beta,\gamma) \approx C_{\gamma-\alpha-\beta+1}.$$

**3. Regularity and moment sequences.** If

$$(13) \qquad \alpha',\beta',\gamma' > 0; \qquad \gamma' > \alpha'+\beta'-1,$$

or if

$$(14) \qquad \alpha',\beta',\gamma' > 0; \qquad \gamma = \alpha+\beta-1,$$

---

[5] H. L. Garabedian and H. S. Wall, "Hausdorff summation and continued fractions," *Transactions of the American Mathematical Society*, 48 (1940), pp. 185-207, p. 196. In this paper, the parameters $\alpha$, $\beta$, $\gamma$ are restricted to positive real numbers for which $\beta$ is an integer and $\gamma > \alpha+\beta-1$.

then $C_{\gamma-\alpha-\beta+1}$ is regular and (11) holds. Hence if (13) or (14) holds then $(H, \alpha, \beta, \gamma)$ is regular and it follows that the generating sequence $\lambda_n(\alpha, \beta, \gamma)$ of hypergeometric coefficients must be a regular moment sequence.

If

$$(15) \qquad \alpha', \beta', \gamma' > 0; \qquad \alpha' + \beta' - 2 < \gamma' < \alpha' + \beta' - 1,$$

or if

$$(16) \qquad \alpha', \beta', \gamma' > 0; \qquad \gamma' = \alpha' + \beta' - 1; \qquad \gamma \neq \alpha + \beta - 1,$$

then (11) holds and the transformations $C_{\gamma-\alpha-\beta+1}$ and $(H, \alpha, \beta, \gamma)$ are not regular and in fact not conservative. It follows that if (15) or (16) holds, then $\lambda_n(\alpha, \beta, \gamma)$ is not a regular moment sequence and in fact is not a moment sequence.

It can be shown easily by another method that if

$$(17) \qquad \alpha, \beta, \gamma \neq 0, -1, -2, \cdots; \qquad \gamma' < \alpha' + \beta' - 1,$$

then $(H, \alpha, \beta, \gamma)$ is not conservative and $\lambda_n(\alpha, \beta, \gamma)$ is not a moment sequence. Under the hypothesis (17), Stirling's formula shows that

$$| \lambda_n(\alpha, \beta, \gamma) | = A n^{\delta}(1 + o_n)$$

where $A$ is a constant not 0, $\delta = -(\gamma' - \alpha' - \beta' - 1) > 0$, and $o_n \to 0$ as $n \to \infty$. The sequence $\lambda_n(\alpha, \beta, \gamma)$ is accordingly unbounded and cannot be a moment sequence, and hence $(H, \alpha, \beta, \gamma)$ is not conservative.

CORNELL UNIVERSITY,
ITHACA, N. Y.

# INFINITE GROUPS GENERATED BY EQUILONG TRANSFORMATIONS OF PERIOD TWO.*[1]

By Edward Kasner and John De Cicco.

**1. Introduction.** The problem of this paper is to determine the infinite groups generated by all equilong transformations of period two. In our preceding work,[2] we found that the set of all equilong transformations of period two may be classified into *three* distinct types: $(T_1)$ equilong involutions, $(T_2)$ $K$ symmetries, and $(T_3)$ $D$ inversions. This is in contrast with the conformal theory where Kasner has proved that the set of all conformal transformations of period two consists of two distinct types: $(\mathcal{T}_1)$ conformal involutions, and $(\mathcal{T}_2)$ conformal symmetries (Schwarzian reflections).

*We thus have five distinct types of transformations of period two: three equilong and two conformal.*

The conformal types and the infinite groups generated by them have been discussed by Kasner.[3] In this paper we shall consider the analogous situations in equilong geometry.

Conformal transformations are correspondences between the $\infty^2$ points of the plane which preserve or reverse the angle between the two directions of any two curves at their common point of intersection. Equilong transformations are correspondences between the $\infty^2$ lines of the plane which preserve or reverse the distance between the two points of contact of any two curves along their common tangent line. Conformal transformations are defined by monogenic functions of the complex variable $x_1 \pm iy_1$, where $i^2 = -1$ and $(x_1, y_1)$ are the cartesian coördinates of a point; whereas equilong transformations are given by monogenic functions of the *dual* variable $x \pm jy$, where $j^2 = 0$ and $(x, y)$ are the hessian or equilong coördinates of a line.

We shall consider only those equilong transformations of the plane which

* Received March 4, 1941.

[1] Presented to the American Mathematical Society, April, 1940.

[2] Kasner and De Cicco, "Equilong and conformal transformations of period two," *Proceedings of the National Academy of Sciences*, vol. 26 (1940), pp. 471-476.

[3] Kasner, "Infinite groups generated by conformal transformations of period two (Involutions and symmetries)," *American Journal of Mathematics*, vol. 38 (1916), pp. 177-184; Kasner, "Conformal geometry," *Proceedings of the Fifth International Congress of Mathematicians*, vol. 2 (1912), p. 81.

convert the positive $y_1$-axis (with equilong coördinates $(0, 0)$) into itself and are regular. Such regular equilong transformations are expressed by power series of the two forms

$$\text{(1)} \qquad \begin{aligned} Z &= \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \cdots, \alpha_1 \alpha_1 \neq 0; \\ \bar{Z} &= \alpha'_1 z + \alpha'_2 z^2 + \alpha'_3 z^3 + \cdots, \alpha'_1 \bar{\alpha}'_1 \neq 0, \end{aligned}$$

where $\bar{z} = x - jy$ is the conjugate of the dual variable $z = x + jy$ and the coefficients are arbitrary dual numbers of the form $a + jb$. Separating (1) into real parts, we find

$$\text{(2)} \qquad \begin{aligned} X &= a_1 x + a_2 x^2 + a_3 x^3 + \cdots, Y = y(a_1 + 2a_2 x + 3a_3 x^2 + \cdots) \\ &\qquad + (b_1 x + b_2 x^2 + b_3 x^3 + \cdots), a_1 \neq 0; \\ X &= a'_1 x + a'_2 x^2 + a'_3 x^3 + \cdots, Y = -y(a'_1 + 2a'_2 x + 3a'_3 x^2 + \cdots) \\ &\qquad - (b'_1 x + b'_2 x^2 + b'_3 x^3 + \cdots), a'_1 \neq 0. \end{aligned}$$

The *direct* equilong transformations, as given by the first of equations (1) or by the first two of equations (2), form a continuous infinite group $G$. The *reverse* equilong transformations, as given by the second of equations (1) or by the last two of equations (2), do *not* form a group. However, if we add this set to our group $G$, we obtain a mixed group $G'$.

We determine all regular equilong transformations of period two. In the direct type $Z = f(z)$ the functional equation is $f[f(z)] \equiv z$, that is $f^2 = 1$; and in the reverse type $\bar{Z} = f(z)$ the functional equation is $\bar{f}[f(z)] \equiv z$, that is $\bar{f}f = 1$ where $\bar{f}$ denotes the series whose coefficients are the conjugates of the coefficients of the series $f$. Thus there are two types of functional equations which are to be solved. The first type yields as a solution a set $(T_1)$ of equilongly equivalent transformations (besides the identical transformation). On the other hand, the second type yields as a solution *two equilongly distinct* sets $(T_2)$ and $(T_3)$.

It can be shown that any equilong involution $(T_1)$ can be reduced to the simple form $Z = -z$ (ordinary symmetry in the positive $y_1$-axis); that any $K$ symmetry $(T_2)$ may be changed to the simple form $Z = \bar{z}$ (ordinary symmetry in the origin accompanied by reversal of orientation); and that any $D$ inversion $(T_3)$ can be reduced to the simple form $Z = -\bar{z}$ (ordinary symmetry in the $x_1$-axis accompanied by reversal of orientation). We observe also that the third type $(T_3)$ is equivalent under the equilong group to Laguerre inversion. Of course, these simple transformations are considered to be correspondences between the lines of the plane.

In our calculations, we find a striking resemblance between the roles played by $(T_1)$ and $(T_3)$ while $(T_2)$ stands apart.

The equilong involutions generate an infinite group $G'_{\mathrm{invol}}$. Any transformation of this group can be factored into involutions in an infinitude of ways, of which at least one will contain *four or fewer* factors. The transformations which are products of an even number of equilong involutions form a continuous subgroup $G_{\mathrm{invol}}$ of $G'_{\mathrm{invol}}$.

Any $K$ symmetry may be defined as the unique reverse equilong transformation which leaves fixed the tangent lines of a given curve $C$.[4] In this respect, $K$ symmetry may be considered to be the dual of conformal symmetry which may be defined as the unique reverse conformal transformation which leaves fixed the points of $C$. The group $K'_{\mathrm{sym}}$ generated by all $K$ symmetries consists of $K$ symmetries and $K$ translations (the products of *two* $K$ symmetries). Any transformation of this group sends any line into one parallel to itself such that the distance between any two parallel lines is preserved or reversed. The $K$ translations form a continuous subgroup $K_{\mathrm{sym}}$ of $K'_{\mathrm{sym}}$.

Finally we shall discuss the group $D'_{\mathrm{invers}}$ generated by all $D$ inversions. Any transformation of this group can be factored into $D$ inversions in an infinitude of ways, of which at least one will contain *four or fewer* factors. If the number of factors is even, the resulting transformations form a continuous subgroup $D_{\mathrm{invers}}$ of $D'_{\mathrm{invers}}$.

Our fundamental result is that *the group generated by all equilong transformations of period two is identical with the group generated by all $K$ symmetries and $D$ inversions.* In particular the groups $G'_{\mathrm{invol}}$, $K'_{\mathrm{sym}}$, and $D'_{\mathrm{invers}}$ are all subgroups of this larger group. See the exact formulation of the final group in Theorem 18.

**2. The discussion of ($T_1$) equilong involutions, ($T_2$) $K$ symmetries, and ($T_3$) $D$ inversions.** First let us note that any direct equilong transformation may be written in either of the two forms (see [2])

$$(3)\quad \begin{aligned} Z - z &= \beta_1(Z+z) + \beta_2(Z+z)^2 + \beta_3(Z+z)^3 + \cdots, \beta_1 \neq \pm 1; \\ Z + z &= \gamma_1(Z-z) + \gamma_2(Z-z)^2 + \gamma_3(Z-z)^3 + \cdots, \gamma_1 \neq \pm 1. \end{aligned}$$

Since an equilong involution is a direct equilong transformation of period two, we obtain from the preceding equations the following result:

THEOREM 1. *Any equilong involution is given in the implicit dual variable form by*

$$(4)\qquad Z + z = \alpha_2(Z-z)^2 + \alpha_4(Z-z)^4 + \alpha_6(Z-z)^6 + \cdots,$$

[4] Kasner, "Equilong symmetry with respect to any curve," *Proceedings of the National Academy of Sciences*, vol. 26 (1940), pp. 287-291.

*where the $\alpha_{2n}$ are arbitrary dual numbers. In the implicit real form, any equilong involution is*

$$(5)\quad \begin{aligned} X + x &= a_2(X-x)^2 + a_4(X-x)^4 + a_6(X-x)^6 + \cdots, \\ Y + y &= (Y-y)[2a_2(X-x) + 4a_4(X-x)^3 + 6a_6(X-x)^5 + \cdots] \\ &\quad + [b_2(X-x)^2 + b_4(X-x)^4 + b_6(X-x)^6 + \cdots]. \end{aligned}$$

*Any equilong involution in the explicit dual variable form is*

$$(6)\qquad Z = -z + \gamma_2 z^2 - \gamma_2^2 z^3 + \gamma_4 z^4 + (-3\gamma_2\gamma_4 + 2\gamma_2^4)z^5 + \cdots,$$

*where the $\gamma_{2n}$ are arbitrary dual numbers. Finally in the explicit real form any equilong involution is*

$$(7)\quad \begin{aligned} X &= -x + \lambda_2 x^2 - \lambda_2^2 x^3 + \lambda_4 x^4 + (-3\lambda_2\lambda_4 + 2\lambda_2^4)x^5 + \cdots, \\ Y &= y[-1 + 2\lambda_2 x - 3\lambda_2^2 x^2 + 4\lambda_4 x^3 + 5(-3\lambda_2\lambda_4 + 2\lambda_2^4)x^4 + \cdots] \\ &\quad + [d_2x^2 - 2\lambda_2 d_2 x^3 + d_4 x^4 + (-3\lambda_2 d_4 - 3\lambda_4 d_2 + 8\lambda_2^3 d_2)x^5 + \cdots]. \end{aligned}$$

In the power series expansion for $X$, the odd coefficients are polynomial functions of the arbitrary even coefficients. In the power series expansion of the part in $Y$ not containing $y$, the odd coefficients are polynomial functions of the even coefficients in this part and of the even coefficients in $X$.

Next let us note that any reverse equilong transformation may be written in either of the two forms

$$(8)\quad \begin{aligned} \bar{Z} - z &= \beta_1(\bar{Z}+z) + \beta_2(\bar{Z}+z)^2 + \beta_3(\bar{Z}+z)^3 + \cdots, \beta_1 \neq \pm 1; \\ \bar{Z} + z &= \gamma_1(\bar{Z}-z) + \gamma_2(\bar{Z}-z)^2 + \gamma_3(\bar{Z}-z)^3 + \cdots, \gamma_1 \neq \pm 1. \end{aligned}$$

Now any $K$ symmetry may be defined as a reverse equilong transformation of period two, which may be written in the form given by the first of the preceding equations. Hence

THEOREM 2. *Any $K$ symmetry is given in the implicit dual variable form by*

$$(9)\qquad \bar{Z} - z = j[b_1(\bar{Z}+z) + b_2(\bar{Z}+z)^2 + b_3(\bar{Z}+z)^3 + \cdots],$$

*where the $b$'s are real numbers. In the real explicit form, any $K$ symmetry may be written as*

$$(10)\qquad X = x, Y = -y + 2(d_1x + d_2x^2 + d_3x^3 + \cdots).$$

*In the explicit dual variable form, any $K$ symmetry is*

$$(11)\qquad \bar{Z} = z - 2j(d_1z + d_2z^2 + d_3z^3 + \cdots).$$

We observe that the $K$ symmetry (10) or (11) may be defined as the unique reverse equilong transformation which leaves fixed the tangent lines of the curve $y = d_1x + d_2x^2 + d_3x^3 + \cdots$.

Finally any $D$ inversion is a reverse equilong transformation of period two, which may be written in the form given by the second of equations (8). Therefore

THEOREM 3. *Any $D$ inversion is given in the implicit dual variable form by*

$$\begin{aligned}(12)\quad \bar{Z} + z = {} & [a_2(\bar{Z} - z)^2 + a_4(\bar{Z} - z)^4 + a_6(\bar{Z} - z)^6 + \cdots] \\ & + j[b_1(\bar{Z} - z) + b_3(\bar{Z} - z)^3 + b_5(\bar{Z} - z)^5 + \cdots],\end{aligned}$$

*where the $a$'s and $b$'s are arbitrary real numbers. In the implicit real form, any $D$ inversion may be written as*

$$\begin{aligned}& X + x = a_2(X - x)^2 + a_4(X - x)^4 + a_6(X - x)^6 + \cdots, \\ (13)\quad & Y - y = (Y + y)\,[2a_2(X - x) + 4a_4(X - x)^3 + 6a_6(X - x)^5 + \cdots] \\ & \qquad - [b_1(X - x) + b_3(X - x)^3 + b_5(X - x)^5 + \cdots].\end{aligned}$$

*In the real explicit form any $D$ inversion is*

$$\begin{aligned}& X = -x + \lambda_2 x^2 - \lambda_2^2 x^3 + \lambda_4 x^4 + (-3\lambda_2\lambda_4 + 2\lambda_2^4)x^5 + \cdots, \\ (14)\quad & Y = -y[-1 + 2\lambda_2 x - 3\lambda_2^2 x^2 + 4\lambda_4 x^3 + 5(-3\lambda_2\lambda_4 + 2\lambda_2^4)x^4 + \cdots] \\ & \qquad + [d_1 x - (3/2)\lambda_2 d_1 x^2 + d_3 x^3 \\ & \qquad + 5/2(-\lambda_2 d_3 - \lambda_4 d_1 + 2\lambda_2^3 d_1)x^4 + \cdots].\end{aligned}$$

In the power series expansion for $X$, the odd coefficients are polynomial functions of the arbitrary even coefficients. In the power series expansion of the part in $Y$ not containing $y$, the even coefficients are polynomial functions of the odd coefficients in this part and of the even coefficients in $X$.

Thus we see that in the $K$ symmetry (10), all the $d$'s, both odd and even, are arbitrary, but in the $D$ inversion (14) the answer depends on the even $\lambda$'s and the odd $d$'s.

**3. The reduction of equilong transformations of period two to canonical forms.** First we shall demonstrate the existence of at least one real function $\Phi(x) = \Sigma\lambda_n x^n$, $(\lambda_1 \neq 0)$, such that $\Phi(X) + \Phi(x) = 0$, where $X$ is given by the first of equations (5). That is, we must have

$$\begin{aligned}(15)\qquad & \lambda_1(X + x) + \lambda_2(X^2 + x^2) + \lambda_3(X^3 + x^3) \\ & \qquad + \lambda_4(X^4 + x^4) + \cdots = 0,\ \lambda_1 \neq 0.\end{aligned}$$

Upon making the substitution $X - x = t$, we find that we can solve this and the first of equations (5) for $x$ and $X$ in terms of $t$ obtaining

$$(16)\quad x = \tfrac{1}{2}(-t + a_2 t^2 + a_4 t^4 + \cdots),\quad X = \tfrac{1}{2}(t + a_2 t^2 + a_4 t^4 + \cdots).$$

Substituting these values of $x$ and $X$ into (15) and equating the coefficients of the different powers of $t$ to zero, we obtain the system of equations

$$
(17)\quad
\begin{aligned}
&\tfrac{1}{2}\lambda_2 + a_2\lambda_1 = 0,\\
&\tfrac{1}{8}\lambda_4 + \tfrac{3}{4}a_2\lambda_3 + \tfrac{1}{2}a_2{}^2\lambda_2 + a_4\lambda_1 = 0,\\
&\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad,\\
&\frac{1}{2^{2n-1}}\lambda_{2n} + \frac{(2n-1)}{2^{2n-2}}a_2\lambda_{2n-1} + L_{2n}(\lambda_1, \lambda_2, \cdots, \lambda_{2n-2}) = 0,\\
&\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad,
\end{aligned}
$$

where $L_{2n}$ is a linear function of $(\lambda_1, \lambda_2, \cdots, \lambda_{2n-2})$ with coefficients which are polynomials in $(a_2, a_4, \cdots, a_{2n})$. This system of equations shows that all the even coefficients $\lambda_{2n}$ are determined as linear functions of odd coefficients $\lambda_{2n-1}$ such that the coefficients of these linear functions are polynomials in the $a_{2n}$. Thus there exist many functions $\Phi(x)$ with the desired property.

Under the direct equilong transformation $X = \Phi(x)$, $Y = y\Phi_x(x)$, our equilong involution (5) becomes $X = -x$, $Y = -y + 2(d_2x^2 + d_4x^4 + \cdots)$. The transform of this under the direct equilong transformation $X = x$, $Y = y - (d_2x^2 + d_4x^4 + \cdots)$ is the equilong involution $X = -x$, $Y = -y$ (or $Z = -z$).

Under the direct equilong transformation $X = x$, $Y = y - (d_1x + d_2x^2 + d_3x^3 + \cdots)$, the $K$ symmetry (10) is reduced to the $K$ symmetry $X = x$, $Y = -y$ (or $Z = \bar{z}$).

Under the direct equilong transformation $X = \Phi(x)$, $Y = y\Phi_x(x)$, our $D$ inversion (13) becomes $X = -x$, $Y = y + 2(d_1x + d_3x^3 + d_5x^5 + \cdots)$. The transform of this under the direct equilong transformation $X = x$, $Y = y + (d_1x + d_3x^3 + d_5x^5 + \cdots)$ is the $D$ inversion $X = -x$, $Y = y$ (or $Z = -\bar{z}$). Therefore we have proved the following theorem:

THEOREM 4. *Any equilong involution can be reduced to the symmetry through the positive $y_1$-axis: $Z = -z$. Any $K$ symmetry may be changed to the symmetry through the origin accompanied by reversal of orientation: $Z = \bar{z}$. Any $D$ inversion may be reduced to the symmetry through the $x_1$-axis accompanied by reversal of orientation: $Z = -\bar{z}$.*

It may be shown without much difficulty that these three sets of equilong transformations of period two are (equilongly) mutually exclusive.

**4. The infinite group $G'_{\text{invol}}$ generated by equilong involutions.** The first question which we wish to consider is that of finding the form of those equilong transformations which can be obtained as the product of two equilong involutions. Given a function $f(z)$ we inquire whether it is possible to find two functions $g(z)$ and $h(z)$ such that, symbolically, $f = hg$, $g^2 = 1$, $h^2 = 1$. If we write the transformations $f$, $g$, $h$ in the real explicit forms

$$
(18) \qquad \begin{aligned} f &: X = f(x),\ Y = yf_x(x) + F(x), \\ g &: X = \phi(x),\ Y = y\phi_x(x) + \psi(x), \\ h &: X = \Phi(x),\ Y = y\Phi_x(x) + \Psi(x), \end{aligned}
$$

our problem is to find the functions $\phi$, $\psi$, $\Phi$, $\Psi$, which satisfy the conditions for equilong involution

$$
(19) \qquad \phi(\phi) = x, \psi(\phi) = -\frac{\psi}{\phi_x}; \quad \Phi(\Phi) = x, \Psi(\Phi) = -\frac{\Psi}{\Phi_x};
$$

and which satisfy the two relations

$$
(20) \qquad \Phi(\phi) = f(x), \quad \psi\Phi_\phi(\phi) + \Psi(\phi) = F(x).
$$

*First we shall proceed to discuss the equation* $\Phi(\phi) = f(x)$. Since $\phi$ and $\Phi$ are involutorial, the leading coefficient of $f(x)$ must necessarily be unity. Our question is that of finding when the equation

$$
(21) \qquad X = x + e_2x^2 + e_3x^3 + e_4x^4 + \cdots,
$$

can result from the elimination of the variable $u$ from the two equations

$$
(22) \qquad \begin{aligned} x + u &= a_2(x-u)^2 + a_4(x-u)^4 + a_6(x-u)^6 + \cdots, \\ u + X &= \lambda_2(u-X)^2 + \lambda_4(u-X)^4 + \lambda_6(u-X)^6 + \cdots, \end{aligned}
$$

defining two arbitrary involutions.

It is convenient to write $e_k = 2^k\epsilon_k$, and to introduce $t = x - u$. Eliminating $x$, $u$, $X$ from this and the preceding three equations, we obtain the following identity in the single variable $t$

$$
(23) \qquad \begin{aligned} -t + P + \epsilon_2P^2 + \epsilon_3P^3 + \epsilon_4P^4 + \cdots &= \lambda_2(t + \epsilon_2P^2 + \epsilon_3P^3 + \epsilon_4P^4 + \cdots)^2 \\ &+ \lambda_4(t + \epsilon_2P^2 + \epsilon_3P^3 + \epsilon_4P^4 + \cdots)^4 \\ &+ \lambda_6(t + \epsilon_2P^2 + \epsilon_3P^3 + \epsilon_4P^4 + \cdots)^6 + \cdots, \end{aligned}
$$

where $P = t + a_2t^2 + a_4t^4 + a_6t^6 + \cdots$. Arranging the preceding equation in powers of $t$ and equating coefficients, we obtain an infinite system of equations which must be discussed.

Let us note that

$$
(24) \qquad \begin{aligned} &\epsilon_2P^2 + \epsilon_3P^3 + \epsilon_4P^4 + \epsilon_5P^5 + \epsilon_6P^6 + \epsilon_7P^7 + \epsilon_8P^8 + \epsilon_9P^9 + \cdots \\ &\quad = \epsilon_2t^2 + (2a_2\epsilon_2 + \epsilon_3)t^3 + (a_2{}^2\epsilon_2 + 3a_2\epsilon_3 + \epsilon_4)t^4 \\ &+ (2a_4\epsilon_2 + 3a_2{}^2\epsilon_3 + 4a_2\epsilon_4 + \epsilon_5)t^5 \\ &+ [2a_2a_4\epsilon_2 + (3a_4 + a_2{}^3)\epsilon_3 + 6a_2{}^2\epsilon_4 + 5a_2\epsilon_5 + \epsilon_6]t^6 \\ &+ [2a_6\epsilon_2 + 6a_2a_4\epsilon_3 + (4a_4 + 4a_2{}^3)\epsilon_4 + 10a_2{}^2\epsilon_5 + 6a_2\epsilon_6 + \epsilon_7]t^7 \\ &+ \left[\begin{aligned} &(2a_2a_6 + a_4{}^2)\epsilon_2 + (3a_6 + 3a_2{}^2a_4)\epsilon_3 + (12a_2a_4 + a_2{}^4)\epsilon_4 \\ &+ (5a_4 + 10a_2{}^3)\epsilon_5 + 15a_2{}^2\epsilon_6 + 7a_2\epsilon_7 + \epsilon_8 \end{aligned}\right]t^8 \\ &+ \left[\begin{aligned} &2a_8\epsilon_2 + (6a_2a_6 + 3a_4{}^2)\epsilon_3 + (4a_6 + 12a_2{}^2a_4)\epsilon_4 + (20a_2a_4 + 5a_2{}^4)\epsilon_5 \\ &+ (6a_4 + 20a_2{}^3)\epsilon_6 + 21a_2{}^2\epsilon_7 + 8a_2\epsilon_8 + \epsilon_9 \end{aligned}\right]t^9 \\ &+ \cdots. \end{aligned}
$$

Substituting this into (23), we see that the equations corresponding to $t^2$ and $t^3$ are

$$a_2 + \epsilon_2 = \lambda_2, \qquad 2\epsilon_2 a_2 + \epsilon_3 = 2\epsilon_2\lambda_2. \tag{25}$$

This gives a necessary relation $\epsilon_3 - 2\epsilon_2^2 = 0$, or $e_3 - e_2^2 = 0$.

Let us assume that $e_2 \neq 0$ so that $\epsilon_2 \neq 0$. The equations corresponding to $t^{2n}$ and $t^{2n+1}$ are of the form

$$\begin{aligned} &a_{2n} - \lambda_{2n} = p_{2n}(a_2, \cdots, a_{2n-2}; \lambda_2, \cdots, \lambda_{2n-2}), \\ &2\epsilon_2(a_{2n} - n\lambda_{2n}) = p_{2n+1}(a_2, \cdots, a_{2n-2}; \lambda_2, \cdots, \lambda_{2n-2}), \end{aligned} \tag{26}$$

where $p_{2n}$ and $p_{2n+1}$ are polynomials in the preceding $a$'s and $\lambda$'s. Therefore, if $n > 1$, we can solve these equations for $a_{2n}$ and $\lambda_{2n}$ as polynomials in the preceding $a$'s and $\lambda$'s. Since for $n = 1$, our equations (25) show only that $\lambda_2 = a_2 + \epsilon_2$ provided that $e_3 - e_2^2 = 0$, it follows that all the coefficients $a_{2n}$ and $\lambda_{2n}$ of our two real involutions (22) can be determined as polynomial functions of $a_2$. Thus

THEOREM 5. *Every real transformation of the form*

$$X = x + e_2x^2 + e_3x^3 + e_4x^4 + \cdots, \quad e_3 - e_2^2 = 0, \qquad e_2 \neq 0, \tag{27}$$

*can be factored into two real involutions. This can be done in $\infty^1$ ways.*

Suppose now that $e_2 = 0 (\epsilon_2 = 0)$. The equations (25) show that $e_3 = 0$ $(\epsilon_3 = 0)$ and $a_2 = \lambda_2$. The equations corresponding to $t^4$, $t^5$, $t^6$, $t^7$, and $t^9$ are

$$\begin{aligned} &a_4 + \epsilon_4 = \lambda_4, \quad 2a_2\epsilon_4 + \epsilon_5 = 0, \quad a_6 - 2a_2^2\epsilon_4 + 3a_2\epsilon_5 + \epsilon_6 = \lambda_6, \\ &4(a_4 - \lambda_4 - 2a_2^3)\epsilon_4 + 4a_2\epsilon_6 + \epsilon_7 = 0, \\ &(4a_6 - 6\lambda_6 - 12a_2^2a_4 - 24a_2^2\lambda_4 - 2a_2^5)\epsilon_4 \\ &\qquad + (10a_2a_4 - 20a_2\lambda_4 - 15a_2^4)\epsilon_5 \\ &\qquad + (6a_4 - 4\lambda_4 - 10a_2^3)\epsilon_6 + 7a_2^2\epsilon_7 + 6a_2\epsilon_8 + \epsilon_9 \\ &\qquad - 8a_2^2\epsilon_4^2 - 2a_2\epsilon_4\epsilon_5 = 0. \end{aligned} \tag{28}$$

The first, second, third, and fifth of these are equivalent to the equations

$$\begin{aligned} &2a_2\epsilon_4 + \epsilon_5 = 0, \quad \lambda_4 = a_4 + \epsilon_4, \quad \lambda_6 = a_6 - 8a_2^2\epsilon_4 + \epsilon_6, \\ &2\epsilon_4 a_6 = -16a_2^2a_4\epsilon_4 + 2a_4\epsilon_6 + 28a_2^5\epsilon_4 - 10a_2^3\epsilon_6 \\ &\qquad + 7a_2^2\epsilon_7 + 6a_2\epsilon_8 + \epsilon_9 - 10\epsilon_4\epsilon_6 + 60a_2^2\epsilon_4^2. \end{aligned} \tag{29}$$

Upon substituting the above values of $a_2$ and $\lambda_4$ in the fourth of equations (28), we obtain the necessary relation

$$4\epsilon_4^4 - \epsilon_5^3 + 2\epsilon_4\epsilon_5\epsilon_6 - \epsilon_4^2\epsilon_7 = 0. \tag{30}$$

Let us now assume that $e_4 \neq 0$ so that $\epsilon_4 \neq 0$. The equations corresponding to $t^{2n}$ and $t^{2n+3}$ are of the form

$$
(31) \qquad \begin{aligned} &a_{2n} - \lambda_{2n} = p_{2n}(a_2, \cdots, a_{2n-2}; \lambda_2, \cdots, \lambda_{2n-2}), \\ &2\epsilon_4(2a_{2n} - n\lambda_{2n}) = p_{2n+3}(a_2, \cdots, a_{2n-2}; \lambda_2, \cdots, \lambda_{2n-2}), \end{aligned}
$$

where $p_{2n}$ and $p_{2n+3}$ are polynomials in the preceding $a$'s and $\lambda$'s. Thus for $n > 3$, we can solve for $a_{2n}$ and $\lambda_{2n}$ as polynomials in the preceding $a$'s and $\lambda$'s. But the equations (29) show that $\lambda_2 = a_2$ is uniquely determined and that $\lambda_4, a_6, \lambda_6$ are determined as polynomials in $a_4$. Therefore all the coefficients $a_{2n}$ and $\lambda_{2n}$ of our two real involutions (22) can be determined as polynomial functions in $a_4$. Thus

THEOREM 6. *Every real function of the form*

$$
(32) \qquad X = x + e_4x^4 + e_5x^5 + e_6x^6 + e_7x^7 + \cdots, \qquad e_4 \neq 0,
$$

*such that*

$$
(33) \qquad 8e_4^4 - e_5^3 + 2e_4e_5e_6 - e_4^2e_7 = 0,
$$

*can be factored into two real involutions. This can be done in $\infty^1$ ways.*

A new type arises when $e_4 = 0$, and so on. The final result is

THEOREM 7. *All real functions which can be obtained as the product of two real involutions are of the form*

$$
(34) \qquad X = x + e_{2k}x^{2k} + e_{2k+1}x^{2k+1} + e_{2k+2}x^{2k+2} + \cdots, \qquad e_{2k} \neq 0,
$$

*where $k = 1, 2, 3, \cdots$, and a single rational relation*

$$
(35) \qquad r_k(e_{2k}, e_{2k+1}, \cdots, e_{4k-1}) = 0,
$$

*holds between the coefficients. This can always be done in $\infty^1$ ways.*

The form of this relation, as well as the number of coefficients involved, changes with the integer $k$, but in all cases $e_{4k-1}$ can be expressed as a rational function of the previous coefficients.

*Next we proceed to discuss the equation* $\psi\Phi_\phi(\phi) + \Psi(\phi) = F(x)$. The problem is to determine the $b_{2n}$ and $d_{2n}$ such that the equation

$$
(36) \qquad \begin{aligned} &[b_2(x-u)^2 + b_4(x-u)^4 + \cdots][-1 + 2\lambda_2(u-X) + 4\lambda_4(u-X)^3 + \cdots] \\ &+ [d_2(u-X)^2 + d_4(u-X)^4 + \cdots][1 + 2a_2(x-u) + 4a_4(x-u)^3 + \cdots] \\ &\quad = [1 + 2a_2(x-u) + 4a_4(x-u)^3 + \cdots] \\ &\quad\quad \times [1 + 2\lambda_2(u-X) + 4\lambda_4(u-X)^3 + \cdots][f_1x + f_2x^2 + f_3x^3 + \cdots], \end{aligned}
$$

where $u$ and $X$ are defined by the equations (21) and (22) is identically satisfied. Let us write $\epsilon_k = 2^ke_k$ and $\eta_k = 2^kf_k$, and introduce the substitution $t = x - u$. Upon eliminating $x$, $u$, $X$ from this and the equations (21), (22), and (36), we obtain the following equation in the single variable $t$

$$
\begin{aligned}
&-[b_2t^2+b_4t^4+\cdots][1+2\lambda_2(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)\\
&+4\lambda_4(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)^3+\cdots]+[1+2a_2t+4a_4t^3+\cdots]\\
&\times[d_2(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)^2+d_4(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)^4+\cdots]\\
&=[1+2a_2t+4a_4t^3+\cdots][1-2\lambda_2(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)\\
&-4\lambda_4(t+\epsilon_2P^2+\epsilon_3P^3+\cdots)^3-\cdots]\\
&\times[\eta_1P+\eta_2P^2+\eta_3P^3+\cdots],
\end{aligned}
\tag{37}
$$

where $P=t+a_2t^2+a_4t^4+\cdots$. On arranging the above equation in powers of $t$ and equating coefficients, we obtain an infinite system of equations which must be discussed.

The three equations corresponding to $t$, $t^2$, and $t^3$ are

$$
\begin{gathered}
\eta_1=0, \qquad -b_2+d_2=\eta_2,\\
-2b_2\lambda_2+2\epsilon_2d_2+2a_2d_2=\eta_3+4a_2\eta_2-2\lambda_2\eta_2.
\end{gathered}
\tag{38}
$$

These equations with (25) give as necessary relations $\eta_1=0$ and $\eta_3-4\epsilon_2\eta_2=0$, or $f_1=0$ and $f_3-2e_2f_2=0$.

Let us assume that $e_2\neq 0$ so that $\epsilon_2\neq 0$. The equations corresponding to $t^{2n}$ and $t^{2n+1}$ are of the form

$$
\begin{gathered}
-b_{2n}+d_{2n}=l_{2n}(b_2,\cdots,b_{2n-2};\ d_2,\cdots,d_{2n-2}),\\
-2\lambda_2b_{2n}+(2n\epsilon_2+2a_2)d_{2n}=l_{2n+1}(b_2,\cdots,b_{2n-2};\ d_2,\cdots,d_{2n-2}),
\end{gathered}
\tag{39}
$$

where $l_{2n}$ and $l_{2n+1}$ are linear polynomials in the preceding $b$'s and $d$'s. Therefore if $n>1$, we can solve these equations for $b_{2n}$ and $d_{2n}$ as linear polynomials in the preceding $b$'s and $d$'s. Since for $n=1$, our equations (38) show only that $d_2=b_2+\eta_2$, we find that all the $b_{2n}$ and $d_{2n}$ can be determined as linear polynomial functions of $b_2$. The coefficients of these are polynomial functions of $a_2$. Hence

THEOREM 8. *Every equilong transformation in the real form*

$$
\begin{aligned}
&X=x+e_2x^2+e_3x^3+\cdots,\\
&Y=y(1+2e_2x+3e_3x^2+\cdots)+(f_2x^2+f_3x^3+\cdots),\\
&e_2\neq 0,\quad e_3-e_2^2=0,\quad f_3-2e_2f_2=0,
\end{aligned}
\tag{40}
$$

*or in the dual variable form*

$$
Z=z+\gamma_2z^2+\gamma_3z^3+\cdots,\ \gamma_2\bar{\gamma}_2\neq 0,\quad \gamma_3-\gamma_2^2=0,
\tag{41}
$$

*can be factored into two equilong involutions. This can be done in $\infty^2$ ways.*

Assume next $e_2=0$. Then $e_3=0$ and $f_3=0$. Thus we have the conditions $\epsilon_2=\epsilon_3=\eta_3=0$. The equations corresponding to $t^2$, $t^4$, $t^5$, $t^6$, and $t^7$ are

$$
(42)\quad
\begin{aligned}
&-b_2+d_2=\eta_2, \quad -b_4+d_4=\eta_4-3a_2^2\eta_2,\\
&\quad 2a_2(-b_4+d_4)-4\lambda_4 b_2+(4a_4+2\epsilon_4)d_2\\
&\quad\quad =\eta_5+4a_2\eta_4+(2a_4-8a_2^3-4\epsilon_4)\eta_2,\\
&-b_6+d_6-2a_2b_2\epsilon_4+(12a_2\epsilon_4+2\epsilon_5)d_2\\
&\quad\quad =\eta_6+5a_2\eta_5+2a_2^2\eta_4+(-14a_2a_4-18a_2\epsilon_4-4a_2^4)\eta_2,\\
&\quad 2a_2(-b_6+d_6)-4\lambda_4 b_4+4(a_4+\epsilon_4)d_4\\
&\quad\quad +b_2(-6\lambda_6-8a_2^2\epsilon_4-2a_2\epsilon_5)+d_2(6a_6+28a_2^2\epsilon_4+14a_2\epsilon_5+2\epsilon_6)\\
&\quad\quad\quad =\eta_7+6a_2\eta_6+6a_2^2\eta_5+(4a_4-12a_2^3-4\epsilon_4)\eta_4\\
&\quad\quad +(8a_6-40a_2^2a_4-36a_2^2\epsilon_4-2a_2\epsilon_5-6\lambda_6)\eta_2.
\end{aligned}
$$

By means of equations (29), the preceding equations are equivalent to

$$
(43)\quad
\begin{aligned}
&d_2=b_2+\eta_2, \quad d_4=b_4+\eta_4-3a_2^2\eta_2,\\
&\quad -2\epsilon_4 b_2=\eta_5+2a_2\eta_4-(2a_4+2a_2^3+6\epsilon_4)\eta_2,\\
&d_6=b_6+\eta_6+8a_2\eta_5+8a_2^2\eta_4-(20a_2a_4+10\,a_2^4+44a_2\epsilon_4)\eta_2,\\
&4\epsilon_6 b_2=-\eta_7-4a_2\eta_6-12a_2^2\eta_5+(-16a_2^3+8\epsilon_4)\eta_4\\
&\quad +(4a_6+32a_2^2a_4+24a_2^5+16a_2^2\epsilon_4+8\epsilon_6)\eta_2.
\end{aligned}
$$

By equations (29) and (43), we obtain the necessary relation

$$
(44)\quad
\begin{aligned}
&2\epsilon_4^5\eta_7-4\epsilon_4^4\epsilon_5\eta_6-4\epsilon_4^4\epsilon_6\eta_5+6\epsilon_4^3\epsilon_5^2\eta_5\\
&\quad +4\epsilon_4^3\epsilon_5\epsilon_6\eta_4-4\epsilon_4^2\epsilon_5^3\eta_4-16\epsilon_4^6\eta_4-4\epsilon_4^4\epsilon_9\eta_2\\
&\quad +12\epsilon_4^3\epsilon_5\epsilon_8\eta_2-7\epsilon_4^2\epsilon_5^2\epsilon_7\eta_2-6\epsilon_4\epsilon_5^3\epsilon_6\eta_2+48\epsilon_4^5\epsilon_6\eta_2\\
&\quad +5\epsilon_5^5\eta_2-68\epsilon_4^4\epsilon_5^2\eta_2=0.
\end{aligned}
$$

Let us now assume $e_4\neq 0$ so that $\epsilon_4\neq 0$. The equations corresponding to $t^{2n}$, $t^{2n+2}$, and $t^{2n+3}$ are of the form

$$
(45)\quad
\begin{aligned}
&-b_{2n}+d_{2n}=l_{2n}(b_2,\cdots,b_{2n-4};\ d_2,\cdots,d_{2n-4}),\\
&-b_{2n+2}+d_{2n+2}=l_{2n+2}(b_2,\cdots,b_{2n-2};\ d_2,\cdots,d_{2n-2}),\\
&2a_2(-b_{2n+2}+d_{2n+2})-4\lambda_4 b_{2n}+(4a_4+2n\epsilon_4)d_{2n}\\
&\qquad\qquad =l_{2n+3}(b_2,\cdots,b_{2n-2};\ d_2,\cdots,d_{2n-2}),
\end{aligned}
$$

where $l_{2n}$, $l_{2n+2}$, and $l_{2n+3}$ are linear polynomials in the preceding $b$'s and $d$'s. These equations show that if $n>2$, the $b_{2n}$ and $d_{2n}$ can be determined as linear polynomials in the preceding $b$'s and $d$'s. This fact combined with equations (43), show that all the $b_{2n}$ and $d_{2n}$ can be determined as linear polynomials in $b_4$ and polynomials in $a_4$. Thus

THEOREM 9. *Every equilong transformation of the form*

$$
(46)\quad
\begin{aligned}
&X=x+e_4x^4+e_5x^5+e_6x^6+e_7x^7+\cdots,\ e_4\neq 0,\\
&Y=y(1+4e_4x^3+5e_5x^4+6e_6x^5+7e_7x^6+\cdots)\\
&\qquad\qquad +(f_2x^2+f_4x^4+f_5x^5+\cdots),
\end{aligned}
$$

*where the two rational relations*

$$
(47)\quad
\begin{aligned}
&8e_4^4 - e_5^3 + 2e_4e_5e_6 - e_4^2e_7 = 0,\\
&2e_4^5f_7 - 4e_4^4e_5f_6 - 4e_4^4e_6f_5 + 6e_4^3e_5^2f_5 + 4e_4^3e_5e_6f_4 - 4e_4^2e_5^3f_4\\
&\quad - 4e_4^6f_4 - 4e_4^4e_9f_2 + 12e_4^3e_5e_8f_2 - 7e_4^2e_5^2e_7f_2 - 6e_4e_5^3e_6f_2\\
&\quad + 24e_4^5e_2f_2 + 5e_5^5f_2 - 34e_4^4e_5^2f_2 = 0,
\end{aligned}
$$

*hold between the coefficients, can be factored into two equilong involutions. This can be done in* $\infty^2$ *ways.*

A new type arises when $e_4 = 0$, and so on. The final result is

THEOREM 10. *All equilong transformations which can be obtained as the products of two equilong involutions are of the form*

$$
(48)\quad
\begin{aligned}
X &= x + e_{2k}x^{2k} + e_{2k+1}x^{2k+1} + e_{2k+2}x^{2k+2} + \cdots, \qquad e_{2k} \neq 0,\\
Y &= y(1 + 2ke_{2k}x^{2k-1} + (2k+1)e_{2k+1}x^{2k}\\
&\quad + (2k+2)e_{2k+2}x^{2k+1} + \cdots) + (f_2x^2 + f_4x^4 + f_5x^5 + \cdots),
\end{aligned}
$$

*where the* $k$ *rational relations*

$$
(49)\quad
\begin{aligned}
&r_k(e_{2k}, e_{2k+1}, \cdots, e_{4k-1}) = 0,\\
&s_k(e_{2k}, e_{2k+1}, \cdots, e_{4k-1};\ f_{2k}, f_{2k+1}, \cdots, f_{4k-1})\\
&\qquad\qquad + t_k(e_{2k}, e_{2k+1}, \cdots, e_{4k+1};\ f_2, f_4, \cdots, f_{2k-2}) = 0,\\
&f_5 = f_5(e_{2k}, e_{2k+1}, e_{2k+2}, e_{2k+3};\ f_2, f_4),\\
&f_7 = f_7(e_{2k}, \cdots, e_{2k+5};\ f_2, f_4, f_6), \cdots,\\
&\qquad\qquad f_{2k-1} = f_{2k-1}(e_{2k}, \cdots, e_{4k-3}, f_2, f_4, \cdots, f_{2k-2}),
\end{aligned}
$$

*hold between the coefficients. This can always be done in* $\infty^2$ *ways.*

The forms of these $k$ relations depend on the integer $k$, but in all cases $e_{4k-1}$ can be expressed as a rational function of the preceding $e$'s, the $f_5, f_7, \cdots, f_{2k-1}$ can be obtained as rational functions of the preceding $e$'s and even $f$'s, and $f_{4k-1}$ can be determined as a rational function of the preceding $e$'s and $f$'s.

By Theorem 10, we find that for each value of the integer $k$, there is a single set of equilong transformations which can be factored into two equilong involutions. *No one of these sets has the group property.* The same is true of the totality of all the sets.

All transformations in question are, however, included in the larger class

$$
(50)\qquad Z = z + \gamma_2z^2 + \gamma_3z^3 + \cdots, \qquad \gamma_3 - \gamma_2^2 = 0,
$$

which *does* constitute a group, as may be verified immediately.

A simple example of a transformation of the group (50), which is not of the form specified in Theorems 8, 9, or 10, and hence cannot be factored into two equilong involutions is $Z = z + z^4$. We shall now show, however,

that *every transformation of the group* (50) *can be factored into four or fewer equilong involutions.*

If, in (50), $\gamma_2$ does not have its real part zero, we already know by Theorem 8 that two equilong involutions are sufficient. Consider, therefore, any transformation $T$ of our group (50) for which the real part of $\gamma_2$ does vanish. Of course, then $\gamma_3 = 0$ on account of the relation $\gamma_3 - \gamma_2^2 = 0$. We can factor $T$ into two transformations $T'$ and $T''$ both of the form (50) and such that the real parts of the coefficients $\gamma_2'$ and $\gamma_2''$ do not vanish. This is seen from the fact that the product $T'T''$ is of the form $z + (\gamma_2' + \gamma_2'')z^2 + \cdots$, and hence the real part of the coefficient of $z^2$ can be made to vanish without taking either of the real parts of $\gamma_2'$ and $\gamma_2''$ equal to zero. We already know that $T'$ and $T''$ can be factored into two equilong involutions. Hence $T$ can be factored into four equilong involutions.

If we multiply (40) or (41) by a general equilong involution, we obtain a type which can be factored into *three* equilong involutions. We thus have completely proved the following proposition.

THEOREM 11. *The only transformations which can be obtained as products of equilong involutions are of the two types*

$$(51) \qquad \begin{aligned} Z &= z + \gamma_2 z^2 + \gamma_3 z^3 + \cdots, \qquad & \gamma_3 - \gamma_2^2 = 0, \\ Z &= -z + \gamma'_2 z^2 + \gamma'_3 z^3 + \cdots, \qquad & \gamma'_3 + \gamma'^2_2 = 0. \end{aligned}$$

*These form a mixed group which we denote by* $G'_{\text{invol}}$. *This is the group generated by all equilong involutions.*

**5. The infinite group $K'_{\text{sym}}$ generated by all $K$ symmetries.** Any $K$ symmetry may be defined as the unique reverse equilong transformation which leaves fixed the tangent lines of a given curve $C$. When this curve $C$ is given, the direct construction of our $K$ symmetry is easy. Let $l$ be any oriented line in the plane. Construct the tangent line $t$ of the curve $C$ which is parallel to $l$. The correspondent $L$ of $l$ under our $K$ symmetry is the line parallel to both $l$ and $t$ such that $t$ is midway between $l$ and $L$. Thus

THEOREM 12. *The construction of $K$ symmetry with respect to the curve $C$ is accomplished by means of ordinary symmetry in the parallel tangent lines of $C$.*

The product of two $K$ symmetries is not a $K$ symmetry. We shall call any such correspondence an *equilong translation.* Concerning the group generated by all $K$ symmetries, we discover the following proposition.

THEOREM 13. *The group generated by all $K$ symmetries is a certain mixed group $K'_{\text{sym}}$ consisting of $K$ symmetries and $K$ translations, expressed as follows*

(52) $$X = x,\ Y = \mp y + (d_1 x + d_2 x^2 + d_3 x^3 + \cdots).$$

*In the dual variable form, this may be written as*

(53) $$\bar{Z} \text{ or } Z = z \mp j(d_1 z + d_2 z^2 + d_3 z^3 + \cdots).$$

*In either form, the first correspondence represents a K symmetry and the second a K translation.*

Any transformaton of this group may be defined as a correspondence which carries any line into one parallel to itself and which preserves or reverses the distance between any two parallel lines. Of course, this group $K'_{\text{sym}}$ contains the continuous group of equilong translations as a subgroup.

**6. The infinite group $D'_{\text{invers}}$ generated by all $D$ inversions.** The next question that we wish to consider is that of finding the form of those equilong transformations which can be obtained as the product of two $D$ inversions. Given a function $Z = f(z)$, we inquire whether it is possible to find two functions $\bar{Z} = g(z)$ and $\bar{Z} = h(z)$ such that symbolically, $f = \bar{h}g$, $\bar{g}g = 1$, $\bar{h}h = 1$. If we write the transformations $f$, $g$, $h$ in the real explicit forms

(54) $$\begin{aligned} f &: X = f(x), \quad Y = yf_x(x) + F(x),\\ g &: X = \phi(x), \quad Y = -y\phi_x(x) + \psi(x),\\ h &: X = \Phi(x), \quad Y = -y\Phi_x(x) + \Psi(x), \end{aligned}$$

our problem is to find the functions $\phi$, $\psi$, $\Phi$, $\Psi$, which satisfy the conditions for $D$ inversion

(55) $$\phi(\phi) = x,\ \psi(\phi) = \frac{\psi}{\phi_x}; \qquad \Phi(\Phi) = x,\ \Psi(\Phi) = \frac{\Psi}{\Phi_x};$$

and which are such that the two relations

(56) $$\Phi(\phi) = f(x),\ -\psi\Phi_\phi(\phi) + \Psi(\phi) = F(x),$$

are identically satisfied.

We have already discussed the equation $\Phi(\phi) = f(x)$ in connection with the corresponding problem about equilong involutions. Therefore *it only remains to discuss the equation* $-\psi\Phi_\phi(\phi) + \Psi(\phi) = F(x)$. The problem is to determine the $b_{2n-1}$ and $d_{2n-1}$ such that the equation

(57) $$\begin{aligned} &-[b_1(x-u) + b_3(x-u)^3 + \cdots]\\ &\quad \times [-1 + 2\lambda_2(u-X) + 4\lambda_4(u-X)^3 + \cdots]\\ &+ [d_1(u-X) + d_3(u-X)^3 + \cdots]\\ &\quad \times [1 + 2a_2(x-u) + 4a_4(x-u)^3 + \cdots]\\ &= [1 + 2a_2(x-u) + 4a_4(x-u)^3 + \cdots]\\ &\quad \times [1 + 2\lambda_2(u-X) + 4\lambda_4(u-X)^3 + \cdots]\\ &\quad \times [f_1 x + f_2 x^2 + f_3 x^3 + \cdots], \end{aligned}$$

where $u$ and $X$ are defined by the equations (21) and (22), is identically satisfied. Let us write $\epsilon_k = 2^k e_k$ and $\eta_k = 2^k f_k$, and introduce the substitution $t = x - u$. Upon eliminating $x$, $u$, $X$ from this and the equations (21), (22) and (57), we obtain the following equation in the single variable $t$

$$
\begin{aligned}
&-[b_1 t + b_3 t^3 + \cdots] \\
&\quad\times [1 + 2\lambda_2(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots) + 4\lambda_4(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots)^3 + \cdots] \\
&+ [1 + 2a_2 t + 4a_4 t^3 + \cdots] \\
&\quad\times [d_1(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots) + d_3(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots)^3 + \cdots] \\
&= [1 + 2a_2 t + 4a_4 t^3 + \cdots] \\
&\quad\times [-1 + 2\lambda_2(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots) + 4\lambda_4(t + \epsilon_2 P^2 + \epsilon_3 P^3 + \cdots)^3 + \cdots] \\
&\quad\times [\eta_1 P + \eta_2 P^2 + \eta_3 P^3 + \cdots],
\end{aligned} \tag{58}
$$

where $P = t + a_2 t^2 + a_4 t^4 + a_6 t^6 + \cdots$. Arranging the above equation in powers of $t$ and equating coefficients, we obtain an infinite system of equations which must be discussed.

The two equations corresponding to $t$ and $t^2$ are

$$
-b_1 + d_1 = -\eta_1, \quad -2\lambda_2 b_1 + d_1(2a_2 + \epsilon_2) = -\eta_2 + \eta_1(-a_2 + 2\epsilon_2). \tag{59}
$$

The equations corresponding to $t^{2n-1}$ and $t^{2n}$ are of the form

$$
\begin{aligned}
&-b_{2n-1} + d_{2n-1} = l_{2n-1}(b_1, \cdots, b_{2n-3};\ d_1, \cdots, d_{2n-3}), \\
&-2\lambda_2 b_{2n-1} + [2a_2 + (2n-1)\epsilon_2] d_{2n-1} \\
&\qquad\qquad = l_{2n}(b_1, \cdots, b_{2n-3};\ d_1, \cdots, d_{2n-3}).
\end{aligned} \tag{60}
$$

The above two sets of equations demonstrate immediately that, if $\epsilon_2 \neq 0$, all the $b_{2n-1}$ and $d_{2n-1}$ are uniquely determined as polynomials in $a_2$.

THEOREM 14. *Every equilong transformation in the real form*

$$
\begin{aligned}
X &= x + e_2 x^2 + e_3 x^3 + e_4 x^4 + \cdots, \qquad e_2 \neq 0,\ e_3 - e_2^2 = 0, \\
Y &= y(1 + 2e_2 x + 3e_3 x^2 + 4e_4 x^3 + \cdots) + (f_1 x + f_2 x^2 + f_3 x^3 + \cdots),
\end{aligned} \tag{61}
$$

*or in the dual variable form*

$$
\begin{aligned}
Z &= (1 + jf_1) z + \gamma_2 z^2 + \gamma_3 z^3 + \cdots, \\
\gamma_n &= e_n + jf_n, \qquad e_2 \neq 0,\ e_3 - e_2^2 = 0,
\end{aligned} \tag{62}
$$

*can be factored into two D inversions. This can always be done in $\infty^1$ ways.*

Assume next $e_2 = 0$ so that $e_3 = 0$, or $\epsilon_2 = \epsilon_3 = 0$. The equations corresponding to $t$, $t^2$, $t^3$, and $t^4$ are

$$
\begin{aligned}
&-b_1 + d_1 = -\eta_1, \quad a_2\eta_1 = \eta_2, \quad -b_3 + d_3 = -\eta_3 + 2a_2^2\eta_1, \\
&2a_2(-b_3 + d_3) - 4\lambda_4 b_1 + (\epsilon_4 + 4a_4) d_1 \\
&\qquad\qquad = -\eta_4 - 3a_2\eta_3 + \eta_1(-a_4 + 4\epsilon_4 + 7a_2^3).
\end{aligned} \tag{63}
$$

The equations corresponding to $t^{2n-1}$, $t^{2n+1}$, and $t^{2n+2}$ are of the form

$$
(64)\quad
\begin{aligned}
&-b_{2n-1}+d_{2n-1}=l_{2n-1}(b_1,\cdots,b_{2n-5};\ d_1,\cdots,d_{2n-5}),\\
&-b_{2n+1}+d_{2n+1}=l_{2n+1}(b_1,\cdots,b_{2n-3};\ d_1,\cdots,d_{2n-3}),\\
&\quad 2a_2(-b_{2n+1}+d_{2n+1})\\
&-4\lambda_4 b_{2n-1}+d_{2n-1}[4a_4+(2n-1)\epsilon_4]=l_{2n+2}(b_1,\cdots,b_{2n-3};\ d_1,\cdots,d_{2n-3}),
\end{aligned}
$$

where $l_{2n-1}$, $l_{2n+1}$, $l_{2n+2}$ are linear polynomials in the preceding $b$'s and $d$'s. These two sets of equations show that all the $b_{2n-1}$ and $d_{2n-1}$ are uniquely determined.

THEOREM 15. *Every equilong transformation of the form*

$$
(65)\quad
\begin{aligned}
&X=x+e_4x^4+e_5x^5+e_6x^6+\cdots,\qquad e_4\neq 0,\\
&Y=y(1+4e_4x^3+5e_5x^4+6e_6x^5+\cdots)\\
&\qquad\qquad+(f_1x+f_2x^2+f_3x^3+\cdots),
\end{aligned}
$$

*where the two rational relations*

$$
(66)\quad 8e_4^4-e_5^3+2e_4e_5e_6-e_4^2e_7=0,\qquad e_5f_1+2e_4f_2=0,
$$

*hold between the coefficients, can be factored into two $D$ inversions. This can always be done in $\infty^1$ ways.*

New types arise when $e_4=0$, and so on. The final result is:

THEOREM 16. *All equilong transformations which can be obtained as the products of two $D$ inversions are of the form*

$$
(67)\quad
\begin{aligned}
&X=x+e_{2k}x^{2k}+e_{2k+1}x^{2k+1}+\cdots,\qquad e_{2k}\neq 0,\\
&Y=y(1+2ke_{2k}x^{2k-1}+(2k+1)e_{2k+1}x^{2k}+\cdots)\\
&\qquad\qquad+(f_1x+f_2x^2+f_3x^3+\cdots),
\end{aligned}
$$

*where the $k$ rational relations*

$$
(68)\quad
\begin{aligned}
&r_k(e_{2k},e_{2k+1},\cdots,e_{4k+1})=0,\quad f_2=f_2(e_{2k},e_{2k+1};f_1),\\
&f_4=f_4(e_{2k},e_{2k+1},e_{2k+2},e_{2k+3};f_1,f_3),\cdots,\\
&f_{2k-2}=f_{2k-2}(e_{2k},\cdots,e_{4k-3};f_1,f_3,\cdots,f_{2k-3}),
\end{aligned}
$$

*hold between the coefficients. This can always be done in $\infty^1$ ways.*

By the preceding theorem there is, for each value of the integer $k$, a set of equilong transformations which can be factored into two $D$ inversions. *No one of these sets has the group property.* The same is true of the totality of all these sets. By an argument similar to the one used in proving Theorem 11, we obtain the following result:

THEOREM 17. *The only transformations which can be obtained as products of $D$ inversions are of the two types*

$$
(69) \qquad \begin{aligned} Z &= (1 + jf_1)z + \gamma_2 z^2 + \gamma_3 z^3 + \cdots, & e_3 - e_2^2 = 0; \\ \bar{Z} &= (-1 + jf'_1)z + \gamma'_2 z^2 + \gamma'_3 z^3 + \cdots, & e'_3 + e'^2_2 = 0, \end{aligned}
$$

*where* $\gamma_n = e_n + jf_n$ and $\gamma'_n = e'_n + jf'_n$. *These form a mixed group* $D'_{\text{invers}}$. *This is the group generated by all D inversions.*

If the group $K'_{\text{sym}}$, given by equations (52), is multiplied by the group $D'_{\text{invers}}$, given by the preceding equations, the following fundamental proposition is obtained:

THEOREM 18. *The group generated by all equilong transformations of period two is*

$$
(70) \qquad \begin{aligned} Z \text{ or } \bar{Z} &= (1 + jf_1)z + \gamma_2 z^2 + \gamma_3 z^3 + \cdots, & e_3 - e_2^2 = 0, \\ Z \text{ or } \bar{Z} &= (-1 + jf'_1)z + \gamma'_2 z^2 + \gamma'_3 z^3 + \cdots, & e'_3 + e'^2_2 = 0, \end{aligned}
$$

*where* $\gamma_n = e_n + jf_n$ and $\gamma'_n = e'_n + jf_n$. *Any transformation of this group may be factored into K symmetries and D inversions in an infinitude of ways, of which at least one will contain either one K symmetry and four or fewer D inversions, or four or fewer D inversions.*

It is observed that the equilong theory as developed in this paper differs very essentially from the corresponding conformal theory. Thus in the equilong theory both the $K$ symmetries and $D$ inversions are needed to obtain the entire group generated by all equilong transformations of period two, whereas in the conformal theory the total group generated by conformal transformations of period two can be factored into conformal symmetries only.

COLUMBIA UNIVERSITY, NEW YORK,
ILLINOIS INSTITUTE OF TECHNOLOGY, CHICAGO.

# UNIVERSAL FUNCTIONS OF EXTENDED POLYGONAL NUMBERS.*

By L. W. Griffiths.

1. **Introduction.** Certain universal functions of extended polygonal numbers were obtained in an earlier paper.[1] In the notations of that paper $m$ is a fixed but arbitrary integer $\geqq 3$. The extended polygonal numbers are the integers

$$(1) \qquad e(x) = -x + m(x^2 + x)/2, \qquad (x = -1, 0, 1, 2, \cdots).$$

All universal functions were determined if $m = 3, 4, 5$, with the sum of the coefficients in a function $\leqq m + 1$. The reason for this condition on the sum of the coefficients was stated in that paper. All functions which might be universal were determined if $m \geqq 6$, with the sum of the coefficients in a function $\leqq m$. Certain of these functions were proved universal. In this paper the remaining functions are proved universal. Thus all universal functions of extended polygonal numbers, with the sum of the coefficients $\leqq m$, have been determined.

2. **The universal functions if $m \geqq 6$.** The following notations are restated from the earlier paper on extended polygonal numbers. The integer $m$ is fixed, whereas $n, a_1, a_2, \cdots, a_n$ are positive integers to be determined. For $k = 1, \cdots, n$ the $e_k$ are extended polygonal numbers $e(x_k)$. The function $f = a_1p_1 + \cdots + a_np_n$ is said to represent the positive integer $A$ when $p_1, \cdots, p_n$ can be chosen so that $a_1p_1 + \cdots + a_np_n = A$; and $f$ is said to be universal when $f$ represents every positive integer. Also

$$1 \leqq a_1 \leqq \cdots \leqq a_n, \quad w_k = a_1 + \cdots + a_k, \quad (a_1, \cdots, a_n) = a_1p_1 + \cdots + a_np_n.$$

It was proved that no function represents the integers $0, 1, 2, \cdots, 3m - 4$ if $w_n < m$. It was also proved that if $w_n = m$ then $f$ represents the integers $0, 1, 2, \cdots, 6m - 3$ if and only if $f$ is one of the following functions:

$$(2) \quad (1, 1, 1, 1, 1, \cdots, a_n), \quad a_k \leqq w_{k-1} - 1 \ (5 \leqq k \leqq n), \quad w_n = m \geqq 6,$$

* Received April 9, 1941.

[1] L. W. Griffiths, "Representation by extended polygonal numbers and by generalized polygonal numbers," *American Journal of Mathematics*, vol. 55 (1933), pp. 102-110.

$$(3)\quad (1, 1, 1, 1, 2, \cdots, a_n), \quad a_k \leqq w_{k-1} - 1 \ (5 \leqq k \leqq n), \quad w_n = m \geqq 6,$$

$$(4)\quad (1, 1, 1, 2, \cdots, a_n), \quad a_k \leqq w_{k-1} - 1 \ (5 \leqq k \leqq n), \quad w_n = m \geqq 6.$$

THEOREM. *If $f$ satisfies (2) or (3) or (4) then $f$ is universal.*

The fact that $f$ is universal if $f$ satisfies (2) was proved in the earlier paper on extended polygonal numbers. In Theorem 2 of that paper it was also proved that if $f$ satisfies (3) then $f$ represents all positive integers $A$ except perhaps those such that $105m - 14 < A < 296m - 80$. In that same theorem it was further proved that if $f$ satisfies (4) then $f$ represents all positive integers $A$ except perhaps those such that $105m - 14 < A < 513m + 210$. It has been proved now that if $f$ satisfies (3) then $f$ represents each integer $A$ such that $105m - 14 < A < 296m - 80$, and that if $f$ satisfies (4) then $f$ represents each integer $A$ such that $105m - 14 < A < 513m + 210$.

The general methods of proof will be illustrated for integers $A$ such that $151m - 22 < A < 153m - 23$. Define $f_2 = (a_3, \cdots, a_n)$, and write $b_i = a_{i+2}$ $(i = 1, \cdots, n-2)$ and $b_1 + \cdots + b_i = w'_i$. Now $f_2$ represents the integers $0, 1, \cdots, w'_{n-2}$, because, as noted at the bottom of page 104 of the earlier paper on extended polygonal numbers, the lemmas 1, 2 of another paper[2] are valid for extended polygonal numbers. Hence $f_2$ represents the integers $0, 1, \cdots, m-2$. But obviously $f = a_1e_1 + a_2e_2 + f_2$ and $a_1 = a_2 = 1$. Hence $f$ represents the integers $e_1 + e_2$, $e_1 + e_2 + 1, \cdots, e_1 + e_2 + m - 2$. If $e_1$ is taken to be the value $e(16) = 136m - 16$ of (1), and if $e_2$ is taken to be the value $e(5) = 15m - 5$, then it is seen that $f$ represents the integers $151m - 21, \cdots, 152m - 23$.

Again, if $f_3$ is defined to be $(a_4, \cdots, a_n)$ and if the notations $c_i = a_{i+3}$ $(i = 1, \cdots, n-3)$ and $w_i'' = c_1 + \cdots + c_i$ are used, then it is seen by Lemma 1 of the paper just mentioned that $f_3$ represents the integers $0, 1, \cdots, w''_{n-3}$ except perhaps $w''_{i-1} + 1 + A''_{i+1}$. Since $c_i - 1 = w''_{i-1} + 1$, $c_i = a_{i+3}$ and $w''_{i-1} = w_{i+2}$, it is true that $f_3$ represents the integers $0, 1, \cdots, m-3$ except perhaps $a_{i+3} - 1 + A_{i+4}$. Hence $f$ represents the integers $e_1 + e_2 + e_3$, $e_1 + e_2 + e_3 + 1, \cdots, e_1 + e_2 + e_3 + m - 3$ except perhaps $e_1 + e_2 + e_3 + a_{i+3} - 1 + A_{i+4}$. If $e_1$ and $e_2$ are taken as before, and if $e_3$ is taken to be the value $e(1) = m - 1$, then it is seen that $f$ represents the integers $152m - 22, \cdots, 153m - 25$ except perhaps $152m - 22 + a_{i+3} - 1 + A_{i+4}$. Now $e(4) = 10m - 4$ and $e(3) = 6m - 3$. Therefore $e(16) + e(4) + e(3) = 152m - 23$. Hence this possible exception

[2] L. W. Griffiths, "A generalization of the Fermat theorem on polygonal numbers," *Annals of Mathematics*, ser. 2, vol. 31 (1930), pp. 1-12.

$152m - 23 + a_{i+3} + A_{i+4}$ is actually represented by $f$ with $e_1 = e(16)$, $e_2 = e(4)$, $e_3 = e(3)$, $e_4 = \cdots = e_{i+2} = e(0) = 0$, $e_{i+3} = e(-1) = 1$, and the remaining $e_j$ either 0 or 1 according as $a_j$ does not or does appear in $A_{i+4}$.

Finally, that $f$ represents $153m - 24$ is proved as follows. Note that $e(15) = 120m - 15$, $e(6) = 21m - 6$, $e(14) = 105m - 14$, $e(8) = 36m - 8$, $e(7) = 28m - 7$. Then if $f$ satisfies (4) with $a_5 = 4$, the equation $153m - 24 = 105m - 14 + 36m - 8 + 2(6m - 3) + 4$ indicates that the choice $e_1 = e(14)$, $e_2 = e(8)$, $e_3 = 0$, $e_4 = e(3)$, $e_5 = 1$, $e_6 = \cdots = e_n = 0$ gives a representation of $153m - 24$. The equation $153m - 24 = 120m - 15 + 21m - 6 + 2(6m - 3) + 3$ is used if $f$ satisfies (4) with $a_5 = 2$ or 3. The equation $153m - 24 = 120m - 15 + 28m - 7 + 5(m - 1) + 3$ is used if $f$ satisfies (3) with $a_6 = 5$. But if $f$ satisfies (3) with $a_6 = 2$, 3, or 4, then the first equation is again used. It remains to represent $153m - 24$ by the function $(1, 1, 1, 1, 2)$ of (3). Here $m = 6$, and therefore $153m - 24 = 154m - 30 = 91m - 13 + 45m - 9 + 15m - 5 + 3(m - 1)$ indicates the representation.

NORTHWESTERN UNIVERSITY.

## RELATIVISTIC EQUATIONS OF MOTION IN ELECTRO-MAGNETIC THEORY.*

By P. R. Wallace.

**Introduction.** It has been found that under certain simplifying assumptions it is possible, on the basis of Maxwell's equations and the relativistic field equations, to obtain a closed form of the equations of motion of charged particles represented as singularities of the field.[1] The significant assumption made in order to obtain this result was that gravitational effects (i. e., those arising from non-linear terms involving the masses of the particles) were negligible in comparison with "purely electromagnetic" effects. In the present paper this assumption is dispensed with and a general formulation of the equations of motion is obtained. It is found in this case that it is no longer possible to express these equations in a closed form, and that recourse must be had to the "new approximation method" which has been used previously to derive the gravitational equations of motion from the relativistic field equations.[2,3] This makes it possible to obtain a series of successive approximations to the equations of motion based essentially on the smallness of the velocities of the particles in comparison with the velocity of light.

We shall attempt here only the problem of two bodies under mutual interaction. Further difficulties arising from the extension to $n$ bodies are formal only, and not fundamental.[4] The equations of motion will be solved up to a specified approximation for two oppositely charged particles, and the results compared with those obtained by others.

The general outline of the method is based on E. I. H. and E. I., and the calculation of the equations of motion follows the lines indicated in these papers except for the parts concerning radiation, which derive from the paper

* Received March 27, 1941.

[1] Infeld and Wallace, *Phys. Rev.*, vol. 57 (1940), p. 797.

[2] Einstein, Infeld and Hoffmann, *Ann. Math.*, vol. 39 (1938), p. 65.

[3] Einstein and Infeld, *Ann. Math.*, vol. 41 (1940), p. 455. The first of these papers will be referred to hereafter as E. I. H., the second as E. I.

[4] This problem has been worked out by the author in a thesis "On the Relativistic Equations of Motion in Electromagnetic Theory," University of Toronto, 1940. This thesis, in which the details of the calculations involved in the present paper may be found, will be referred to as T.

on gravitational radiation by Infeld.[5] For the sake of brevity and convenience, specific references will be made to all these papers in the exposition to follow.

The calculations involved in the solution of the equations of the two-body problem are very similar to those of Robertson[6] in the gravitational problem.

**1. Derivation of the equations of motion.** We shall start from field equations which differ from those of E. I. H. by the inclusion of the electromagnetic energy-momentum tensor, i. e.,

$$R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R + 8\pi\kappa T_{\mu\nu} = 0, \qquad (\mu, \nu = 0, 1, 2, 3), \tag{1.1}$$

in the usual notation, $\kappa$ being the gravitational constant. $T_{\mu\nu}$ is defined by:

$$T_{\mu\nu} = -[g_{\mu\lambda}F^{\lambda\rho}F_{\nu\rho} - \tfrac{1}{4}g_{\mu\nu}F^{\rho\sigma}F_{\rho\sigma}]/4\pi. \tag{1.2}$$

The velocity of light will be taken as unity throughout. The electromagnetic field tensor $F_{\nu\rho}$ may be expressed in terms of the four-potential $\gamma_\nu$:

$$F_{\nu\rho} = \gamma_{\nu;\rho} - \gamma_{\rho;\nu} = \gamma_{\nu|\rho} - \gamma_{\rho|\nu}, \tag{1.3}$$

where the notation "$_{;\rho}$" indicates covariant differentiation with respect to $x^\rho$, the coördinates of space-time, and the stroke indicates ordinary differentiation with respect to a coördinate.

To the equations (1.1) are added the Maxwell equations

$$F^{\mu\rho}{}_{;\rho} = 0. \tag{1.4}$$

These may be written also in terms of the four-potential:

$$g^{\rho\sigma}\gamma_{\mu;\rho\sigma} - \gamma^{\rho}{}_{;\rho\mu} + R_{\mu\rho}\gamma^{\rho} = 0. \tag{1.5}$$

As is customary in solving these equations, we shall assume $\gamma^{\rho}{}_{;\rho} = 0$, so that the second term of (1.5) vanishes.

(1.1) and (1.5) constitute a set of ten independent equations in fourteen unknowns. As in E. I., the addition of four non-tensorial "coördinate conditions" gives the problem the correct degree of determinacy, and gives rise to equations of motion of singularities of the field (i. e., of $g_{\mu\nu}$ and $\gamma_\nu$), which occur as conditions for the consistency of the set of equations consisting of field equations and coördinate conditions.

The existence of equations of motion follows as in E. I. If we define

$$S_{\mu\nu} = T_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\eta^{\rho\sigma}T_{\rho\sigma}, \tag{1.6}$$

where $\eta_{\mu\nu}$ is the metric of empty space-time, the equations (1.4), (1.5) and (1.6) of the paper E. I. will be altered by the addition to their left-hand sides

---

[5] Infeld, *Phys. Rev.*, vol. 53 (1938), p. 836. To be referred to as I.

[6] Robertson, *Ann. Math.*, vol. 39 (1938), p. 104.

of $16\pi\kappa(S_{mn}, S_{00}, S_{0m})$ respectively, $(m, n = 1, 2, 3)$. Therefore the equations of motion will become

$$\int^{k} (2\Lambda_{mn} + 16\pi\kappa S_{mn}) \cos(x^n, N)\, dS = 0, \tag{1.7}$$

$$\int^{k} (2\Lambda_{0n} + 16\pi\kappa S_{0n}) \cos(x^n, N)\, dS = 0, \tag{1.8}$$

where the "$k$" above the integral sign indicates that the surface encloses the $k$-th singularity. Because the divergence of the integrand vanishes in each case, the integrals (1. 7) and (1. 8) do not depend on the shape of the surface of integration. They depend only on the coördinates of the singularities and their time derivatives.

The approximation method used is that of E. I. (§ 4). The expansion of the field quantities in terms of the parameter $\lambda$ makes it possible to solve the equations (1. 1) and (1. 5) along with the coördinate conditions by the method of successive approximations. As in E. I. also, the components of the metric tensor are replaced as unknowns by the quantities $\gamma_{\mu\nu}$ defined there, and the coördinate conditions chosen are

$$\gamma_{0n|n} - \gamma_{00|0} = 0, \tag{1.9}$$

$$\gamma_{mn|n} = 0. \tag{1.10}$$

In the first approximation these are invariant under a Galileian transformation.

From the solutions of these equations we may calculate successive approximations to the equations of motion. These equations may be written

$$\sum_{l=l_0}^{\infty} \lambda^l \overset{k}{\underset{l}{C}}_m = 0, \tag{1.11}$$

$$\sum_{l=l'_0}^{\infty} \lambda^l \overset{k}{\underset{l}{C}}_0 = 0, \tag{1.12}$$

where

$$4\pi \overset{k}{\underset{l}{C}}_m = \int^{k} (2\underset{l}{\Lambda}_{mn} + 16\pi\kappa \underset{l}{S}_{mn}) \cos(x^n, N)\, dS$$

and

$$4\pi(\overset{k}{\underset{l}{C}}_0 - \tfrac{1}{3}\overset{k}{\underset{l}{C}}_s \dot{\overset{k}{\xi}}{}^s) = \int^{k} (2\underset{l}{\Lambda}_{0n} + 16\pi\kappa \underset{l}{S}_{0n}) \cos(x^n, N)\, dS.$$

The $\overset{k}{\xi}{}^s$ are the coördinates of the $k$-th singularity, and dots represent derivatives with respect to the "auxiliary" time $\tau = \lambda x^0$. The numbers written beneath the various quantities indicate the power of $\lambda$ with which each is associated in the power series expansion in which it occurs.

The calculation of the gravitational potentials $\gamma_{\mu\nu}$ is similar to that of E. I. H., these potentials being modified by the addition of terms arising from the inclusion of the quantities $S_{\mu\nu}$ in the field equations. The electromagnetic potentials $\gamma_\nu$ are calculated as in I., being modified only by the addition of "interaction" terms (i. e., terms involving products of mass and charge) arising from the entry of the gravitational potentials into the equations (1.5).

As in Infeld and Wallace [1], the possibility arises of choosing one of two possible forms of solution for the potentials. One proceeds in alternate powers of $\lambda$ only, and may be identified as the "standing wave" solution. This solution does not take account of the effect of radiation. The other, which proceeds in consecutive powers, is the solution generally known as the "retarded potential," and does take account of radiation. From the formal point of view they are equally valid, though physical considerations give preference to the latter.

The details of the calculations of the equations of motion may be found in T.; we shall be content here merely to state the results. It may be noted that the fourth equation of motion (1.8) imposes no restriction on the motion of the singularities in addition to those imposed by (1.7).

The first non-zero terms in the expansions of $\overset{k}{C}_m$ are $\underset{4}{\overset{k}{C}}_m$, $(k=1,2)$, and these have the values

$$(1.13)\qquad \begin{aligned} \underset{4}{\overset{1}{C}}_m &= 4\kappa\left\{ m_1\ddot{\eta}^m - \kappa m_1 m_2 \frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) + e_1 e_2 \frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right)\right\}, \\ \underset{4}{\overset{2}{C}}_m &= 4\kappa\left\{ m_2\ddot{\zeta}^m - \kappa m_1 m_2 \frac{\partial}{\partial\zeta^m}\left(\frac{1}{r}\right) + e_1 e_2 \frac{\partial}{\partial\zeta^m}\left(\frac{1}{r}\right)\right\}, \end{aligned}$$

where $r=[(\eta^s-\zeta^s)(\eta^s-\zeta^s)]^{1/2}$; $\eta^s, \zeta^s$ are the space coördinates of the two particles, $m_1, m_2$ are their masses and $e_1, e_2$ are their charges, respectively. On equating the expressions for $\underset{4}{\overset{k}{C}}_m$ to zero, we obtain a first approximation to the equations of motion. Thus the motion in the lowest approximation is governed by an inverse square law, which is a combination of Newton's law of gravitation and Coulomb's electrostatic law for the interaction of charges.

In calculating the subsequent approximations, these equations of motion may be used to simplify the integrals obtained.

The terms $\underset{5}{\overset{k}{C}}_m$ are identically zero, so that the next approximation is that which involves the terms $\underset{6}{\overset{k}{C}}_m$. These may be split into three parts:

$$(1.14)\qquad \underset{6}{\overset{k}{C}}_m = \underset{6}{\overset{k}{C}}_m{}' + \underset{6}{\overset{k}{C}}_m{}'' + \underset{6}{\overset{k}{C}}_m{}''',$$

each of which bears to the one preceding a ratio of the order of magnitude of $\kappa m^2/e^2$. We may call the three terms the electromagnetic, interaction and gravitational terms respectively. The latter are given explicitly in E. I. H., equation (16. 1); in the units of this paper the expression given there would be multiplied by $\kappa^3$. The electromagnetic and interaction parts are, for the first particle:

$$\overset{1}{\underset{6}{C}}{}_m{}' = -4\kappa e_1 e_2 \Big[ \frac{1}{2} \frac{\partial}{\partial \eta^m}\left(\frac{1}{r}\right) \dot{\eta}^s \dot{\eta}^s + \frac{\partial}{\partial \eta^s}\left(\frac{1}{r}\right) \dot{\eta}^s \dot{\eta}^m + \frac{\partial}{\partial \eta^m}\left(\frac{1}{r}\right) \dot{\eta}^s \dot{\zeta}^s + \frac{\partial}{\partial \eta^s}\left(\frac{1}{r}\right) \dot{\eta}^s \dot{\zeta}^m - \frac{1}{2} \frac{\partial^3 r}{\partial \eta^m \partial \eta^r \partial \eta^s} \dot{\zeta}^r \dot{\zeta}^s + \frac{\partial}{\partial \eta^s}\left(\frac{1}{r}\right) \dot{\zeta}^s \dot{\zeta}^m + \frac{e_1 e_2}{m_2} \frac{1}{r} \frac{\partial}{\partial \zeta^m}\left(\frac{1}{r}\right) \Big], \tag{1.15}$$

and

$$\overset{1}{\underset{6}{C}}{}_m{}'' = -4\kappa^2 e_1 e_2 (4m_1 + 4m_2) \frac{1}{r} \frac{\partial}{\partial \eta^m}\left(\frac{1}{r}\right) \tag{1.16}$$

respectively. (These are calculated in T.)

Up to this approximation the results obtained are the same whether the standing wave or retarded potential solutions are used. It is when we come to calculate $\underset{7}{C}_m$ that the difference in the solutions makes itself shown. For if the standing wave solutions are used, $\underset{7}{C}_m = 0$, but the retarded potential solutions yield

$$\overset{1}{\underset{7}{C}}{}_m = -(8/3)\kappa e_1 (e_1 \dddot{\eta}^m + e_2 \dddot{\zeta}^m). \tag{1.17}$$

This term may be called the "radiation term," since it arises only when we take account of radiation in the calculation of the potentials.

**2. Solution of the equations of motion.** The equations of motion of the first particle may be written

$$\lambda^4 \overset{1}{\underset{4}{C}}{}_m + \lambda^6 \overset{1}{\underset{6}{C}}{}_m + \lambda^7 \overset{1}{\underset{7}{C}}{}_m = 0$$

to the approximation considered (see (1. 11)). If now we alter the units by going back to the original time coördinate $x^0$ instead of $\tau$ and change the units of mass and charge from $m$ and $e$ to $M$ and $E$, where $M = \lambda^2 m$ and $E = \lambda^2 e$, we may write the equations in a form not involving $\lambda$ explicitly. It will not cause confusion if we keep the notation for time, mass and charge used throughout, and understand the equations to be written in terms of the new units.

With this in mind, we may write the equations of motion

$$
\begin{aligned}
&m_1\ddot{\eta}^m + e_1e_2\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) = e_1e_2[(\tfrac{1}{2}\dot{\eta}^s\dot{\eta}^s + \dot{\eta}^s\dot{\zeta}^s)\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right)\\
&\qquad + (\dot{\eta}^s\dot{\eta}^m - \dot{\eta}^s\dot{\zeta}^m + \dot{\zeta}^s\dot{\zeta}^m)\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right) - \frac{1}{2}\frac{\partial^3 r}{\partial\eta^m\partial\eta^r\partial\eta^s}\dot{\zeta}^r\dot{\zeta}^s]\\
&\qquad - \frac{e_1{}^2e_2{}^2}{m_2}\frac{1}{r}\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) + \frac{2}{3}e_1(e_1\dddot{\eta}^m + e_2\dddot{\zeta}^m),
\end{aligned}
\tag{2.1 (a)}
$$

and a similar equation in which the masses, charges and coördinates of the particles are interchanged, which may be designated as (2. 1) (b). In these equations, $\underset{6}{\overset{1}{C}}{}_m''$ and $\underset{6}{\overset{1}{C}}{}_m'''$ are neglected in comparison with $\underset{6}{\overset{1}{C}}{}_m'$.

We shall solve first the equations of motion without radiation, that is to say, the equations in which the terms of (2. 1) (a) and (b) containing third derivatives are omitted. The effect of radiation will be studied in a separate section.

Let us introduce the variables

$$
(2.2)\qquad
\begin{aligned}
&\text{(a)}\quad \alpha_m = (m_1\eta^m + m_2\zeta^m)/(m_1+m_2),\\
&\text{(b)}\quad \beta_m = \eta^m - \zeta^m.
\end{aligned}
$$

$\alpha_m$ are then the coördinates of the classical "center of mass" of the particles, and $\beta_m$ their relative coördinates. In terms of these variables the equations of motion may be written [7]

$$
\begin{aligned}
&\text{(a)}\quad M\ddot{\alpha}_m = \tfrac{1}{2}e_1e_2(\dot{\eta}^s\dot{\eta}^s - \dot{\zeta}^s\dot{\zeta}^s)\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) + e_1e_2(\dot{\eta}^m\dot{\zeta}^s - \dot{\eta}^s\dot{\zeta}^m)\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right)\\
&\qquad + \tfrac{1}{2}e_1e_2(\dot{\eta}^r\dot{\eta}^s - \dot{\zeta}^r\dot{\zeta}^s)\frac{\partial^3 r}{\partial\eta^m\partial\eta^r\partial\eta^s} + e_1{}^2e_2{}^2\left(\frac{1}{m_1} - \frac{1}{m_2}\right)\frac{1}{r}\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right)\\
&\text{(b)}\quad m_1m_2\ddot{\beta}_m + Me_1e_2\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) = \tfrac{1}{2}e_1e_2(m_2\dot{\eta}^s\dot{\eta}^s + m_1\dot{\zeta}^s\dot{\zeta}^s)\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right)\\
&\qquad + e_1e_2M\dot{\eta}^s\dot{\zeta}^s\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) + e_1e_2(m_2\dot{\eta}^s\dot{\eta}^m + m_1\dot{\zeta}^s\dot{\zeta}^m)\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right)\\
&\qquad - e_1e_2(m_1\dot{\eta}^s\dot{\zeta}^m + m_2\dot{\eta}^m\dot{\zeta}^s)\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right)\\
&\qquad\qquad + e_1e_2(m_1\dot{\eta}^s\dot{\eta}^m + m_2\dot{\zeta}^s\dot{\zeta}^m)\frac{\partial}{\partial\eta^s}\left(\frac{1}{r}\right)\\
&\qquad - \tfrac{1}{2}e_1e_2(m_1\dot{\eta}^r\dot{\eta}^s + m_2\dot{\zeta}^r\dot{\zeta}^s)\frac{\partial^3 r}{\partial\eta^m\partial\eta^r\partial\eta^s} - 2e_1{}^2e_2{}^2\frac{1}{r}\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right),
\end{aligned}
\tag{2.3}
$$

where $M = m_1 + m_2$.

Since $r = (\beta_s\beta_s)^{1/2}$ it follows that $\frac{\partial}{\partial\eta^m}\left(\frac{1}{r}\right) = -\frac{\beta_m}{r^3}$. Therefore, as a first approximation to these equations we may write

[7] Cf. Darwin, *Phil. Mag.*, vol. 39 (1920), p. 537.

$$(2.4)\qquad \begin{aligned} &(a)\quad M\ddot{\alpha}_m = 0, \\ &(b)\quad \ddot{\beta}_m - \frac{Me_1e_2}{m_1m_2}\frac{\beta_m}{r^3} = 0. \end{aligned}$$

In the case where the charges $e_1$ and $e_2$ have opposite signs, these equations have the well-known solutions representing the elliptical orbit,

$$\alpha_m = \bar{\alpha}_m = 0;\quad \beta_1 = \bar{\beta}_1 = \bar{r}\cos\theta;\quad \beta_2 = \bar{\beta}_2 = \bar{r}\sin\theta;$$
$$\beta_3 = \bar{\beta}_3 = 0;\quad u = 1/\bar{r} = (1/p)[1 + \epsilon\cos(\theta - \omega)],$$

where the bars indicate that the values given refer to the first approximation to the motion. The "angular momentum" and "energy" integrals are

$$(2.5)\qquad L = \frac{m_1m_2}{M}\bar{r}^2\dot{\theta} = \left(-\frac{e_1e_2m_1m_2}{M}p\right)^{1/2},$$

$$(2.6)\qquad E = \frac{m_1m_2}{2M}\bar{v}^2 + \frac{e_1e_2}{\bar{r}} = \frac{e_1e_2}{2a}.$$

These equations enable us to obtain the relative velocity $\bar{v}$ and its radial component $\dot{\bar{r}}$ as functions of $u$.

The equations (2.2) may be solved for $\eta^m$ and $\zeta^m$ in terms of $\alpha_m$ and $\beta_m$. In the first approximation, since $\alpha_m = 0$,

$$(2.7)\qquad \eta^m = (m_2/M)\bar{\beta}_m,\qquad \zeta^m = -(m_1/M)\bar{\beta}_m,$$

and these values may be used on the right-hand side of (2.3) (a) and (b).

From (2.5) and (2.6) we deduce that

$$(2.8)\qquad \dot{\bar{r}}^2 = -\frac{Me_1e_2}{am_1m_2}(2au - 1 - apu^2).$$

Furthermore $\bar{\beta}_s\dot{\bar{\beta}}_s = -\dot{u}/u^3 = \dot{\bar{r}}/u$. From the foregoing facts and the equations (2.3) (a) and (b) we deduce the equation

$$(2.9)\qquad \ddot{\alpha}_m = A\bar{\beta}_m + B\dot{\bar{\beta}}_m,$$

where

$$A = -\frac{e_1^2e_2^2\delta m}{2aMm_1m_2}u^3(3apu^2 - 4au + 1);\quad B = -\frac{e_1e_2\delta m}{M^2}\dot{u},\ \text{and}\ \delta m = m_1 - m_2.$$

It will be noted that in the case of the circular orbit $1/u = p = a$, $\dot{u} = 0$, so that $\ddot{\alpha}_m = 0$ and the center of gravity experiences no acceleration whatsoever. This is also true in the case where the masses are equal, since then $\delta m = 0$.

Let us investigate the motion of the center of mass. The secular perturbation of the velocity $\dot{\alpha}_m$ is given by $(1/T)\int_0^T \ddot{\alpha}_m dt$, where $T$ is the period

in the elliptical orbit. If this is written as an integral with respect to $\theta$ and if we substitute from (2. 9), it may be expressed in the form

$$\frac{1}{2\pi a^{3/2}p^{3/2}}\left[\int_0^{2\pi} A\bar{\beta}_m\bar{r}^2 d\theta + \int_0^{2\pi} B\dot{\bar{\beta}}_m\bar{r}^2 d\theta\right].$$

For each value of the index $m$ it is easily shown that the sum of these integrals is zero.

It may therefore be concluded that there is no secular perturbation of the center of mass.

We turn now to a discussion of the relative motion. Making use of the equations (2. 6) and (2. 7) we obtain for $\beta_m$ the equation

$$\ddot{\beta}_m - \frac{Me_1e_2}{m_1m_2}\frac{\beta_m}{r^3} = C\bar{\beta}_m + D\dot{\bar{\beta}}_m, \tag{2.10}$$

where

$$C = \frac{M^2e_1{}^2e_2{}^2}{2am_1{}^2m_2{}^2}u^3([0,3]apu^2 + [1,-1]2au - [1,1])$$

and

$$D = \frac{Me_1e_2}{m_1m_2}[1,0]\dot{u},$$

the square bracket being defined by

$$M^2[k,l] = km_1{}^2 - lm_1m_2 + km_2{}^2.$$

On the right-hand side $\beta_m$ has been replaced by its first approximation $\bar{\beta}_m$, the error introduced in this way being of the eighth order and therefore negligible.

It is obvious that in this approximation, as in the first, $\beta_3$ may be taken equal to zero, and the motion remains plane.

We may introduce the quantities

$$\Lambda = \frac{m_1m_2}{M}(\beta_1\dot{\beta}_2 - \beta_2\dot{\beta}_1), \tag{2.11}$$

$$\mathcal{E} = \frac{m_1m_2}{2M}(\dot{\beta}_1{}^2 + \dot{\beta}_2{}^2) + \frac{e_1e_2}{r}. \tag{2.12}$$

The former is the classical angular momentum and the latter the classical energy of the system relative to the mass center. To the required order of approximation we may write

$$\frac{d\Lambda}{dt} = DL$$

and

$$\frac{d\mathcal{E}}{dt} = \frac{m_1m_2}{M}(C\bar{r}\dot{\bar{r}} + D\bar{v}^2),$$

where $L$ is the angular momentum in the elliptical orbit. The first of these equations may be integrated to give

$$\Lambda = L\left(1 + \frac{Me_1e_2}{m_1m_2}[1,0]u\right). \tag{2.13}$$

Also, substituting in the expression for $d\mathcal{E}/dt$ it is found that

$$\frac{d\mathcal{E}}{dt} = -\frac{Me_1{}^2e_2{}^2}{2am_1m_2}([0,3]apu^2 + [3,-1]2au - [3,1])\dot{u},$$

or, integrating,

$$\mathcal{E} = E\{1 - \frac{Me_1e_2}{m_1m_2}u([0,1]apu^2 + [3,-1]au - [3,1])\}. \tag{2.14}$$

Making use of the fact that

$$\mathcal{E} = \frac{M}{2m_1m_2}\Lambda^2\left[u'^2 + \left(\frac{du'}{d\theta}\right)^2\right] + e_1e_2u',$$

where $u' = 1/r$ in the perturbed orbit, we may deduce an equation in $u'$:

$$\frac{d^2u'}{d\theta^2} + u' = \frac{1}{p}\{1 - \frac{Me_1e_2}{2am_1m_2}([1,1]2au + [1,1] - [0,3]apu^2)\}. \tag{2.15}$$

Since now $u$ is known in terms of $\theta$ the equation can be integrated to this order of approximation. The latter bracketed terms on the right give rise to the perturbation of the elliptical orbit, and are responsible for an advance of perihelion.

Let us investigate the secular change in the longitude of perihelion. The right-hand side of (2.15) may be written as a sum of constants and multiples of $\cos(\theta - \omega)$ and $\cos 2(\theta - \omega)$. The first type of term involves only a constant change in $p$, and the last a periodic perturbation. Omitting these terms, we may solve and obtain

$$u' = \frac{1}{p}\{1 + \epsilon\cos(\theta - \omega) - \frac{Me_1e_2\epsilon}{2m_1m_2p}(\theta - \omega)\sin(\theta - \omega)\}. \tag{2.16}$$

The apsides are found by solving the equation

$$\frac{du'}{d\theta} = -\frac{\epsilon}{p}\sin(\theta - \omega) - \frac{Me_1e_2\epsilon}{2m_1m_2p^2}[\sin(\theta - \omega) + (\theta - \omega)\cos(\theta - \omega)] = 0.$$

One apse is at $\theta = \omega$; the next is at $\theta = \omega + 2\pi + \delta\omega$, where $\delta\omega$ is given by the equation

$$\delta\omega = -\frac{\pi Me_1e_2}{m_1m_2p}.$$

If this is applied to the problem of the hydrogenic atom, $e_1 = -e$,

$e_2 = Ze$, $m_1/m_2 \approx 1/1850$. Changing to natural units by replacing $m$ by $m/c^2$ and $e$ by $e/c^2$, we obtain

$$\delta\omega \approx \frac{\pi Z e^2}{m_1 p c^2}. \tag{2.17}$$

This is precisely the result obtained from other considerations by Sommerfeld.[8]

The above solution applies to the case in which gravitational effects are negligible in comparison with electromagnetic ones. The converse limiting case is the one treated by Einstein, Infeld and Hoffmann.[2] If the problem is attempted in its most general form, proceeding from the complete equations of motion involving electromagnetic, interaction and gravitational terms, no new fundamental difficulties are encountered. The results obtained are a combination of those derived in the two limiting problems. It is found once again that there is no secular perturbation of the mass center. The formula above for the advance of perihelion is modified by the addition of the term $(6\pi\kappa M)/pc^2$, which is the value obtained by Robertson[6] in the gravitational problem.

**3. Effect of radiation on the orbit.** The effect of using the retarded potential solutions for the gravitational and electromagnetic potentials is to give rise to the term $2e_1(e_1\dddot{\eta}^m + e_2\dddot{\zeta}^m)/3$ on the right-hand side of (2.1). This leads to an alteration on the right-hand side of (2.3) (a) of amount $2(e_1+e_2)(e_1\dddot{\eta}^m + e_2\dddot{\zeta}^m)/3$ and a corresponding change in (2.3) (b) of amount $2(m_2e_1 - m_1e_2)(e_1\dddot{\eta}^m + e_2\dddot{\zeta}^m)/3$. Therefore the equation (2.9) for $\alpha_m$ is replaced by

$$\ddot{\alpha}_m = (A + A')\bar{\beta}_m + (B + B')\dot{\bar{\beta}}_m, \tag{3.1}$$

where $A' = 3gu^2\dot{u}$ and $B' = gu^3$, $g$ being defined by

$$g = \frac{e_1e_2(e_1+e_2)}{M}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right).$$

$A$ and $B$ are as in (2.9).

It is easily verified that

$$\int_0^{2\pi}(A'\bar{\beta}_m + B'\dot{\bar{\beta}}_m)\bar{r}^2 d\theta = 0,$$

so that, taking account of radiation to this order, there is still no secular perturbation of the velocity of the center of mass.

---

[8] Sommerfeld, *Atomic Structure and Spectral Lines*, Methuen (1934), pp. 253-4.

The revised equation for $\beta_m$ is

$$\ddot{\beta}_m - \frac{Me_1e_2}{m_1m_2}\frac{\beta_m}{r^3} = (C + C')\bar{\beta}_m + (D + D')\dot{\bar{\beta}}_m, \tag{3.2}$$

where $C$ and $D$ are defined as in (2.10), $C' = 3hu^2\dot{u}$ and $D' = hu^3$, $h$ being a contraction for the factor $\frac{2}{3}\left(\frac{e_1}{m_1} - \frac{e_2}{m_2}\right)^2 e_1e_2$. Calculating the time derivatives of $\Lambda$ and $\mathcal{E}$ and using (3.2) to simplify the result,

$$\frac{d\Lambda}{dt} = DL, + hu^3L, \tag{3.3}$$

$$\frac{d\mathcal{E}}{dt} = \frac{m_1m_2}{M}(C\dot{r}\ddot{r} + D\bar{v}^2), - \frac{he_1e_2}{a}u^3(3apu^2 - 4au + 2), \tag{3.4}$$

where the part of an expression following the comma arises from the radiation terms. The perturbation of $\Lambda$ and $\mathcal{E}$ arising from the terms preceding the commas are periodic and not secular. Hence, *stable elliptical orbits are possible if radiation terms are omitted.*

Let us consider then the effects of radiation. Since $h$ is negative when the charges have opposite signs, it follows that $\Lambda$ decreases monotonically. Furthermore, the radiation term in (3.4) is responsible for a secular decrease in $\mathcal{E}$. The amount of this decrease is found, on calculation, to be $\frac{\pi hL}{p^3}\left(3 - \frac{p}{a}\right)$ per revolution. It must be concluded that *when radiation terms are taken into account, stable elliptical orbits are no longer possible.*

It can be shown that the effect of the last term in (3.3) is to cause a spiralling in of the orbit. For, omitting terms which merely introduce a periodic perturbation in $\Lambda$, we have

$$\frac{d\Lambda}{dt} = \frac{hm_1m_2}{M}u\dot{\theta},$$

and since $u = (1/p)[1 + \epsilon\cos(\theta - \omega)]$ we deduce that

$$\Lambda = \Lambda_0 + \frac{m_1m_2h}{Mp}[\theta - \omega) + \epsilon\sin(\theta - \omega)], \tag{3.5}$$

$\Lambda_0$ being the value of $\Lambda$ given by (2.13). But $\Lambda$ is proportional to the rate at which area is swept out by the line joining the particles. All the terms of the right-hand side of (3.5) are periodic except $\frac{m_1m_2h}{Mp}(\theta - \omega)$, which decreases steadily with time. This indicates a secular decrease of area at a correspondingly increasing rate. In units referred to which the velocity of light is $c$, this rate is $|h(\theta - \omega)/2pc^3|$ (the factor $1/c^3$ may be checked by dimensional considerations).

In the case of circular orbits, the equations (3. 3) and (3. 4) become

$$\frac{d\Lambda}{dt} = \frac{h}{a^3} L = -\sigma, \tag{3.6}$$

$$\frac{d\mathcal{E}}{dt} = -\frac{he_1e_2}{a^4} = -\chi, \tag{3.7}$$

where $\sigma$ and $\chi$ are positive constants. Hence both $\Lambda$ and $\mathcal{E}$ decrease uniformly with time unless radiation terms are omitted, in which case they are constants of the motion.

The accelerations of the two particles (electron and proton) are $f_1 = e_1e_2/m_1a^2$ and $f_2 = e_1e_2/m_2a^2$ respectively in the circular orbits. Therefore, again using units in which the velocity of light is $c$, we may write

$$\frac{d\mathcal{E}}{dt} = -\frac{2}{3}\frac{e_1{}^2f_1{}^2}{c^3} - \frac{2}{3}\frac{e_2{}^2f_2{}^2}{c^3} + \frac{4}{3}\frac{e_1e_2f_1f_2}{c^3}. \tag{3.8}$$

The first two terms are in accord with the familiar expression for the radiation of accelerating electrons. The last term may be compared with the result obtained by Synge[9] for the rate of decrease of energy of a system consisting of several charges. He does not obtain terms corresponding to the first two of (3. 8) because of the difference in his choice of energy-momentum tensor, which is defined in such a way as to involve only interactions between the various particles.

We wish to acknowledge the encouragement and help of Professor Infeld, under whose direction the work appearing in this paper was instituted.

UNIVERSITY OF CINCINNATI.

[9] Synge, *Trans. Roy. Soc. Can.*, Series III, vol. 34 (1940).

# POSTULATES FOR THE THEORY OF PROBABILITY.*[1]

By Arthur H. Copeland.

In order to obtain a clear understanding of how a mathematical theory is to be applied, it is advisable that there be mathematical counterparts of the important activities in which one must engage in applying the theory. In the application of the theory of probability there are two important types of activity: the observation of successes and failures and the assignment of numbers (i. e., probabilities) which are appropriate to such observations. The system of postulates which is presented in this paper differs from other systems for the theory of probability in that it contains a mathematical counterpart of the observation. Without this our ideas concerning the relation between probabilities and observations are likely to remain hazy with regard to such questions as the following. What probabilities are appropriate to given observations? To what extent are these probabilities arbitrary and to what extent are they determined? In what sense can a probability be verified or fail to be verified by observations?[2]

The postulate system of this paper is based on Boolean algebra. The existence of a mathematical counterpart of the observation is made possible by the following devices. First we introduce a subset of the Boolean elements. The elements of this subset are called atomic since they admit of no decomposition within the system. Second, in addition to the operators $\cdot$, $\vee$, $\sim$ (and, or, not) of classical Boolean algebra we shall introduce a new operator $\subset$ (if).

The analogies between the operators $+$, $\times$, $-$ of ordinary algebra and the operators $\vee$, $\cdot$, $\cdot\sim$ (or, and, and not) of Boolean algebra have often been noted. The operator $\div$ has no analogue in the classical Boolean algebra. This analogue is now furnished by the operator $\subset$. Thus the operation $x \div y$ is defined if and only if the denominator $y$ is not equal to zero. Similarly $x \subset y$ is defined if and only if $y$ is not equal to zero. Corresponding to the relation

$$(x \div z) + (y \div z) = (x + y) \div z,$$

* Received February 17, 1941.

[1] Presented to the Society April, 1937.

[2] For a more complete treatment of the philosophical aspects of the theory of probability the reader is referred to Copeland I. References to the literature are given at the end of this article. For other postulational bases for the theory of probability see Koopman, I, II, III, and Evans and Kleene, I.

we have

$$(x \subset z) \vee (y \subset z) = (x \vee y) \subset z.$$

The analogy to the relation

$$(x \div y) \times (z \div u) = (x \times z) \div (y \times u)$$

is not so complete. In Boolean algebra two fractions are multiplied in this simple manner only when they have a common denominator. Thus

$$(x \subset z) \cdot (y \subset z) = (x \cdot y) \subset (z \cdot z) = (x \cdot y) \subset z.$$

Corresponding to the relation

$$x \div y = (x \times z) \div (y \times z),$$

we have

$$x \subset y = (x \cdot y) \subset (y \cdot y) = (x \cdot y) \subset y.$$

Unfortunately the analogy fails at the most important point. The operator $\subset$ does not furnish us with an inverse to multiplication. Thus $y \cdot (x \subset y)$ is not in general equal to $x$. However we do have the equation

$$p(y) \times p(x \subset y) = p(x)$$

whenever $x \cdot y = x$. In this equation $p(y)$, $p(x \subset y)$, and $p(x)$ are respectively the probabilities of $y$, $x \subset y$, and $x$.

The operation $x \subset y$ is similar to the operation $y \supset x$ ($y$ implies $x$) defined in *Principia Mathematica.*[3] However the definition of implication given in *Principia* is not sufficiently general to enable one to handle conditional probabilities or inverse probabilities. The operation $x \subset y$ defined by our system is adequate for such probabilities. In addition it is equivalent to $y \supset x$ in all cases where implication is required in symbolic logic.

It is significant that although Boolean algebra is an important contribution to mathematics, it has been applied only to a very limited extent in that field for which it was intended—namely, the theory of probability. Perhaps the reason for this neglect is that the symbolism is inadequate to handle conditional and inverse probabilities. Now that this deficiency is remedied we obtain an extremely simple and convenient symbolism for the treatment of all probability problems.

Before presenting the postulates I shall make a few remarks concerning notation. In order to avoid a multiplicity of parentheses, brackets, and braces I shall follow the conventions laid down in *Principia* as to the use of dots for bracketing. When a single dot appears in a formula, it is understood that

[3] Whitehead and Russell I.

everything to the right of it is to be treated as though it were enclosed in one bracket and everything to the left of it is to be treated as though it were enclosed in another bracket. Thus $x \vee y \cdot \subset z$ means $(x \vee y)(\subset z)$ or simply $(x \vee y) \subset z$. When further bracketing is required, two dots are used to indicate a stronger bracketing than one, etc. Thus $x \vee y \cdot \subset z : \subset u$ means $[(x \vee y) \subset z] \subset u$. When a dot is used to mean " and," it also has a bracketing force. Hence $(x \cdot y) \subset y$ is written $x \cdot y : \subset y$. The double dot is also used to mean " and " when a stronger bracketing force is needed. Thus $x \cdot \sim (y \vee z)$ is written $x : \sim \cdot y \vee z$. We shall depart from the convention of *Principia* to the extent that the equality sign will always bracket.

**1. The postulates.** We shall be concerned with a set $B$ (Boolean elements), a sub-set $A$ (atoms), and a set $N$ (real numbers). The elements of $B$ can be combined by means of the operators $\cdot$, $\vee$, $\sim$, $\subset$ (and, or, not, if) to form new elements of $B$. An additional operator $p$ (probability of) transforms an element of $B$ into an element of $N$. The expression $p(x)$ means the probability of $x$. We shall now present the postulates. Each one is followed by an interpretation.

POSTULATE 1. *The set $B$ constitutes a Boolean algebra.*

We shall let $U$ denote the unit element and 0 the zero element of the Boolean system. This first postulate is really a number of postulates. See Huntington I and Bernstein I.

POSTULATE 2. *If $x$ is an element of $B$ such that $x \cdot y = x$* or 0 *for every element $y$ of $B$, then $x$ is an element of $A$.*

POSTULATE 3. *If $x$ is any element of $A$ and $y$ is any element of $B$, then $x \cdot y = x$ or 0.*

The atoms are those elements which admit of no decomposition within the system. Hence an atom lies either entirely within or entirely outside of any other element of the system.

POSTULATE 4. *If $x$ and $y$ are two elements of $B$ such that $x \cdot z = y \cdot z$ for every element $z$ of $A$, then $x = y$.*

Two elements $x$ and $y$ are equal if they are composed of the same atoms.

POSTULATE 5. *If $x$ and $y$ are elements of $B$ and $y \neq 0$, then $x \subset y$ is an element of $B$.*

If $x$ and $y$ are events, then " $x$ if $y$ " is an event.

POSTULATE 6. $x \subset y = x \cdot y : \subset y$ if $y \neq 0$.

The event "$x$ if $y$" has the same meaning as the event "$x$ and $y$ if $y$."

POSTULATE 7. *$p(x)$ is an element of $N$.*

The probability of an event is a number. However the expression $p(x)$ may be given an alternative interpretation. In fact it is proved (see Theorem 3) that if $x$ is an atom, then $p(y \subset x)$ is either 1 or 0. The expression $p(y \subset x)$ is then interpreted as an observation. The value 1 designates a success; the value 0, a failure. It is further proved (see Theorem 7) that if $x$ is a disjunction of $n$ atoms $x_1, x_2, \cdots, x_n$, then $p(y \subset x)$ is the success ratio corresponding to the observations $p(y \subset x_1), p(y \subset x_2), \cdots, p(y \subset x_n)$. All of the above interpretations require that when the operator $p$ is applied to a Boolean element, it produces a number. It will be observed that this postulate also consists of a number of postulates since we are demanding that the set of elements $p(x)$ possess the properties of the numerical continuum. See Huntington II.

POSTULATE 8. *If $z$ is an element of $B$ and $x$ and $y$ are elements of $A$ such that $x \cdot z = x \neq 0$ and $y \cdot z = y \neq 0$, then $p(x \subset z) = p(y \subset z)$.*

Roughly this postulate states that all atoms are on a par with respect to probabilities. That is, if $x$ and $y$ are two non-zero atoms each of which is included within an element $z$, then the probability of $x$ on the assumption that $z$ has occurred is the same as the probability of $y$ on the assumption that $z$ has occurred. However the important application of this postulate is in the development of Theorem 7 where $z$ consists of a finite disjunction of atoms and hence $p(x \subset z)$ and $p(y \subset z)$ are interpreted as success ratios instead of probabilities. Therefore it would be preferable to say that atoms are on a par with respect to success ratios.

POSTULATE 9. $p(U \subset x) = 1$ if $x \neq 0$.

POSTULATE 10. $p(0 \subset x) = 0$ if $x \neq 0$.

These postulates state that the probability of a certainty on the assumption that an event $x$ has occurred is 1 and that the probability of an impossibility on the assumption that an event $x$ has occurred is 0.

POSTULATE 11.

$$p(x \vee y \cdot \subset z) = p(x \subset z) + p(y \subset z) - p(x \cdot y : \subset z) \text{ if } z \neq 0.$$

This is the additive law of probability. It can be put into more familiar

form by specializing. First it follows from Postulate 10 that if the events $x$ and $y$ are mutually exclusive (i. e., $x \cdot y = 0$), we obtain the equation

$$p(x \vee y \cdot \subset z) = p(x \subset z) + p(y \subset z).$$

Next it can be proved (see Theorem 10) that if we then substitute $U$ for $z$, we obtain $p(x \vee y) = p(x) + p(y)$. This is the conventional form of the additive law.

POSTULATE 12. *The set $A$ is an $\omega$-series and if 0 is a member of $A$, then 0 is the first element of the series.*

DEFINITION. *An element $x$ of $B$ is called a fundamental segment with respect to an element $n$ of $A$ provided*

(a) $m \cdot x = m$ *for every element $m$ of $A$ which either precedes or is equal to $n$.*

(b) $m \cdot x = 0$ *for every element $m$ of $A$ which follows $n$. The fundamental segment $x$ of the element $n$ is denoted by $(n)$.*

POSTULATE 13. *Given a positive number $\epsilon$ there exists an element $N$ of $A$ such that*

$$| p[x \subset (n)] - p(x) | < \epsilon$$

*for every element $n$ of $A$ which follows $N$.*

Without some such demands as those contained in Postulates 12 and 13 we should be unable to obtain any relation between observations and probabilities. In fact it would in general be true that we could obtain for a given event any one of the success ratios $0, 1/n, 2/n, \cdots, 1$ by a proper choice of $n$ observations.

Postulate 12 demands that the set of atoms shall have the order type of the non-negative integers. Hence it is convenient to use the integers themselves to denote these elements. It is proved (see Theorem 5) that the element 0 actually is atomic and hence must be the first element. Thus the atoms are denoted as follows: $0, 1, 2, \cdots, n, \cdots$.

A fundamental segment $(n)$ can be interpreted as the disjunction of all atoms up to and including $n$. See Theorem 8. The expression $p[x \subset (n)]$ is therefore interpreted as the success ratio for the first $n$ observations of the event $x$. Postulate 13 demands that the limit of this success ratio shall be equal to the probability of the event $x$. Thus Postulates 12 and 13 give us a relationship between observations and probabilities. They demand that this relationship shall exist only for a specified manner of choosing the observations. This requirement does not prevent us from forming success ratios

corresponding to sub-sequences of these observations. In fact the reader is referred to Definition 18 for the consideration of such sub-sequences. The extent to which Postulate 12 restricts the Boolean system will be discussed after Definition 12.

Postulate 12 together with some of the preceding postulates demands that the elements possess the following structure. If $x$ is any element of $B$, then $p(x \subset 1)$, $p(x \subset 2), \cdots$ is a 1, 0-sequence and represents the sequence of observations of $x$. This sequence is called the spectrum of $x$. A spectrum uniquely characterizes an element. That is, two Boolean elements having the same spectrum are equal. See Theorem 9.

POSTULATE 14. *If $x$ and $y$ are elements of $B$, $m$ and $n$ are non-zero elements of $A$, $p[(m) \subset (n)] = p[y \subset (n)]$, and $m$ does not follow $n$, then $x \subset :y \cdot (n) = x \subset y \cdot \subset (m)$.*

POSTULATE 15. *If $x$, $y$, $z$ are elements of $B$, $m$ and $\lambda$ are elements of $A$, $y \neq 0, m \neq 0, \lambda \neq 0, x \subset y = z, (m) \cdot y = y, p[(\lambda) \subset (m)] = p[y \subset (m)]$ and $\lambda$ does not follow $m$, then $z \subset n = z \subset \cdot n + \lambda$ for every element $n$ of $A$ with the exception of* 0.

These postulates prescribe the following relation between the spectra:

$$x^{(1)}, x^{(2)}, \cdots, x^{(k)}, \cdots$$
$$y^{(1)}, y^{(2)}, \cdots, y^{(k)}, \cdots$$
$$z^{(1)}, z^{(2)}, \cdots, z^{(k)}, \cdots$$

of the elements $x$, $y$, and $z = x \subset y$. See Theorem 19. From the spectrum of $y$ let us select in order those terms which are equal to 1 and let us denote them by $y^{(n_1)}, y^{(n_2)}, \cdots, y^{(n_k)}, \cdots$. Then $z^{(k)} = x^{(n_k)}$ where $k = 1, 2, \cdots$. It will be observed that the spectrum of $x \subset y$ is obtained by selecting a sub-sequence of the terms of the spectrum of $x$. The selection operator is $y$. A term of the spectrum of $x$ is selected *if* and only if the corresponding term of the spectrum of $y$ is a 1, i.e., a success. Hence $x \subset y$ appropriately represents the event "$x$ if $y$."

We shall now consider more specifically what is demanded by Postulate 14. If $p[(m) \subset (n)] = p[y \subset (n)]$ and if $m$ does not follow $n$, then $m$ is the number of 1's in the first $n$ terms of the spectrum of $y$. The spectra of $y$ and $y \cdot (n)$ agree in the first $n$ terms and hence the spectra of $x \subset y$ and $x \subset : y \cdot (n)$ agree in the first $m$ terms. Moreover the first $m$ terms of the spectrum of $x \subset y$ are the same as the first $m$ terms of the spectrum of $x \subset y \cdot \subset (m)$. Therefore we have the equality of the first $m$ terms of the spectra of $x \subset : y \cdot (n)$ and $x \subset y \cdot \subset (m)$. We shall see that it is a consequence

of Postulate 15 that for each of these sequences the remainder of the terms is a periodic repetition of the first $m$ terms. Thus $x \subset : y \cdot (n) = x \subset y \cdot \subset (m)$.

Postulate 15 is concerned with an element $y$ whose spectrum contains only a finite number of successes. Such an element is characterized by the fact that there exists an $m$ such that $(m) \cdot y = y$. The conditions that $\lambda$ does not follow $m$ and that $p[(\lambda) \subset (m)] = p[y \subset (m)]$ then tell us that $\lambda$ is the total number of 1's in the spectrum of $y$. The equation $z \subset n = z \subset \cdot n + \lambda$ (where $z = x \subset y$) states that the spectrum of $z$ is periodic and has the period $\lambda$.

**2. The development of the theory of probability.** In this section we shall develop a sufficient number of theorems to assure the reader that the theory of probability can be derived from these postulates. Each theorem will be followed by an indication in parenthesis of the postulates and theorems upon which it is based.

We observe that $p(x \subset x) = p(U \cdot x : \subset x) = p(U \subset x) = 1$. Hence

THEOREM 1. *$p(x \subset x) = 1$ if $x$ is an element of $B$ and $x \neq 0$.*

$(P: 1, 5, 6, 7, 9)$

If $x_1 \cdot x_2 = 0$, then

$$p(x_1 \vee x_2 \cdot \subset y) = p(x_1 \subset y) + p(x_2 \subset y) - p(0 \subset y)$$
$$= p(x_1 \subset y) + p(x_2 \subset y).$$

Hence we obtain inductively

THEOREM 2. $p(x_1 \vee x_2 \vee \cdots \vee x_n \cdot \subset y)$
$$= p(x_1 \subset y) + p(x_2 \subset y) + \cdots + p(x_n \subset y)$$

*if $x_1, x_2, \cdots, x_n$ are mutually exclusive elements of $B$* (i.e., $x_i \cdot x_j = 0$ *whenever $i \neq j$) and if $y$ is an element of $B$ such that $y \neq 0$.*

$(P: 1, 5, 7, 10, 11)$

If $x$ is an element of $A$ and $y$ is an element of $B$, then

$$p(y \subset x) = p(x \cdot y : \subset x) = p(x \subset x) \text{ or } p(0 \subset x).$$

But $p(x \subset x) = 1$ and $p(0 \subset x) = 0$. Hence

THEOREM 3. *If $x$ is an element of $A$ distinct from the zero element and $y$ is an element of $B$, then $p(y \subset x) = 1$ or $0$ according as $x \cdot y = x$* or 0.

$(P: 1, 3, 5, 6, 7, 10. \ T: 1)$

The expression $p(y \subset x)$, where $x$ is an element of $A$ and $y$ is an element of $B$, is interpreted as an observation. We say that a success has been observed when this expression is equal to 1 and a failure when it is equal to 0. An

alternative interpretation is the following. We may regard $x$ as a point and $y$ as a set of points. The expression $p(y \subset x)$ is then the characteristic function of point sets. It has the value 1 when the point $x$ belongs to the set $y$ and the value 0 otherwise.

If $x$ and $y$ are elements of $A$ such that $x \cdot y \neq 0$, then $x \cdot y = x$. Also $x \cdot y = y \cdot x = y$. Hence $x = y$. Therefore

THEOREM 4. *If $x$ and $y$ are elements of $A$ and $x \neq y$, then $x \cdot y = 0$.* $(P: 1, 3)$

The element 0 has the property that if $y$ is any element of $B$, then $0 \cdot y = 0$. Hence

THEOREM 5. *0 is an element of $A$.* $(P: 1, 2)$

If $x_1, x_2, \cdots, x_n$ are distinct elements of $A$, then

$$\begin{aligned} 1 &= p(x_1 \vee x_2 \vee \cdots \vee x_n \cdot \subset \cdot x_1 \vee x_2 \vee \cdots \vee x_n) \\ &= p(x_1 \subset \cdot x_1 \vee x_2 \vee \cdots \vee x_n) \\ &\quad + p(x_2 \subset \cdot x_1 \vee \cdots \vee x_n) + \cdots + p(x_n \subset \cdot x_1 \vee \cdots \vee x_n). \end{aligned}$$

But $x_i \cdot x_1 \vee x_2 \vee \cdots \vee x_n = x_i$ and hence

$$p(x_i \subset \cdot x_1 \vee \cdots \vee x_n) = p(x_j \subset \cdot x_1 \vee \cdots \vee x_n).$$

Therefore

THEOREM 6. *If $x_1, x_2, \cdots, x_n$ are distinct non-zero elements of $A$, then*

$$p(x_i \subset \cdot x_1 \vee x_2 \vee \cdots \vee x_n) = 1/n.$$

$(P: 1, 8.\ T: 1, 2, 4)$

If $x_1, x_2, \cdots, x_n$ are distinct elements of $A$ and $y$ is an element of $B$, then $y \cdot x_i = x_i$ or 0. In either case

$$p(y \cdot x_i : \subset \cdot x_1 \vee x_2 \vee \cdots \vee x_n) = p(y \subset x_i)/n.$$

Hence it follows from Theorem 2 that

THEOREM 7. *If $x_1, x_2, \cdots, x_n$ are distinct non-zero elements of $A$ and $y$ is an element of $B$, then*

$$p(y \subset \cdot x_1 \vee x_2 \vee \cdots \vee x_n) = \sum_{i=1}^{n} p(y \subset x_i)/n.$$

$(P: 1, 3, 5, 6, 7, 10.\ T: 2, 3, 4, 6)$

Let $0, 1, 2, \cdots, n$ be the first $n + 1$ elements of $A$ and let $x = 1 \vee 2 \vee \cdots \vee n$. Then $x \cdot m = m$ if $m$ does not follow $n$ and $x \cdot m = 0$ if $m$ does follow $n$. Therefore

THEOREM 8. *If* $0, 1, 2, \cdots, n$ *are the first* $n+1$ *elements of* $A$, *then*

$$(n) = 1 \vee 2 \vee \cdots \vee n. \qquad (P: 1, 4, 12.\ T: 4)$$

If $x$ and $y$ are elements of $B$ and $n$ is an element of $A$, then $p(x \subset n)$ is equal to 1 or 0 according as $x \cdot n$ is equal to $n$ or 0. Similarly for $y$. Hence $p(x \subset n) = p(y \subset n)$ if $x \cdot n = y \cdot n$. Therefore $x = y$ if $p(x \subset n) = p(y \subset n)$ for every element $n$ of $A$ other than 0. With the aid of Theorem 7 it can be proved inductively that $x = y$ if $p[x \subset (n)] = p[y \subset (n)]$ for every $n$ of $A$ other than 0. Thus

THEOREM 9. *If* $x$ *and* $y$ *are elements of* $B$ *having the same spectrum, then* $x = y$. *Moreover if* $p[x \subset (n)] = p[y \subset (n)]$ *for every* $n$ *of* $A$ *other than* 0, *then* $x = y$. $(P: 1, 4.\ T: 3, 7, 8)$

We observe that $p[(n) \subset (n)] = p[U \subset (n)]$. Thus it follows from Postulate 14 that

$$x \subset (n) = x \subset : U \cdot (n) = x \subset U \cdot \subset (n).$$

Hence

$$p[x \subset (n)] = p[x \subset U \cdot \subset (n)].$$

As a consequence of Theorem 9 we obtain

THEOREM 10. $x \subset U = x$. $(P: 1, 5, 7, 9, 14.\ T: 1, 9)$

The following four theorems are now obtained by simple substitution.

THEOREM 11. $p(x \vee y) = p(x) + p(y) - p(x \cdot y)$. $(P: 1, 11.\ T: 10)$

THEOREM 12. $p(U) = 1$ *and* $p(0) = 0$. $(P: 1, 9, 10.\ T: 10)$

Note that $p(x) = 1$ *does not imply* $x = U$ *and* $p(x) = 0$ *does not imply* $x = 0$.

THEOREM 13. *If* $x_1, x_2, \cdots, x_n$ *are mutually exclusive elements of* $B$, *then*

$$p(x_1 \vee x_2 \vee \cdots \vee x_n) = p(x_1) + p(x_2) + \cdots + p(x_n). \quad (P: 1.\ T: 2, 10)$$

THEOREM 14. $p(x \cdot \sim y) = p(x) - p(x \cdot y)$ *and* $p(\sim x) = 1 - p(x)$. $(P: 1.\ T: 12, 13)$

It follows from Theorems 7 and 8 that $0 \leqq p[x \subset (n)] \leqq 1$. Hence by Postulate 13 we obtain

THEOREM 15. $0 \leqq p(x) \leqq 1$. $(P: 13.\ T: 3, 7, 8)$

If $x$ and $y$ are elements of $B$, then $x \cdot (n) = x_1 \vee x_2 \vee \cdots \vee x_m$ where

$x_1, x_2, \cdots, x_m$ are those non-zero atoms (taken in order) such that $x_i \cdot x = x_i$. Then $p[x \subset (n)] = p[(m) \subset (n)] = m/n$ and, if $x \neq 0$, we can choose $n$ so large that $m \neq 0$. It follows from Theorem 7 and Postulate 14 that

$$\text{(a)} \quad p[y \subset : x \cdot (n)] = \sum_{i=1}^{m} p(y \subset x_i)/m = p[y \subset x \cdot \subset (m)].$$

On the other hand

$$\text{(b)} \quad p[x \cdot y : \subset (n)] = \sum_{k=1}^{n} p(x \cdot y : \subset k)/n = \sum_{i=1}^{m} p(y \subset x_i)/n.$$

Since $p(y \cdot x : \subset k) = 0$ if $k$ is not equal to some $x_i$ and

$$p(y \cdot x : \subset k) = p(y \subset x_i) \text{ if } k = x_i.$$

Comparing (a) and (b) we see that

$$\text{(c)} \quad \begin{aligned} p[x \cdot y : \subset (n)] &= p[y \subset x \cdot \subset (m)] \cdot m/n \\ &= p[y \subset x \cdot \subset (m)] \cdot p[x \subset (n)]. \end{aligned}$$

From this point on we shall consider separately two cases. In the first case $m$ increases without bound as $n$ becomes infinite and in the second case $m$ remains finite. In the first case we obtain

$$\lim_{n \to \infty} p[x \cdot y : \subset (n)] = \lim_{n \to \infty} p[x \subset (n)] \cdot \lim_{m \to \infty} p[y \subset x \cdot \subset (m)]$$

and hence $p(x \cdot y) = p(x) \cdot p(y \subset x)$. In the second case $p(x)$ and $p(y \cdot x)$ are both zero and hence irrespective of the value of $p(y \subset x)$ we have $p(y \cdot x) = p(x) \cdot p(y \subset x)$. Therefore

THEOREM 16. $p(x \cdot y) = p(x) \cdot p(y \subset x)$ *if* $x \neq 0$.
($P$: 1, 3, 4, 5, 6, 7, 12, 13, 14. $T$: 3, 7, 8)

By interchanging $x$ and $y$ in Theorem 16 we obtain

$$p(x \cdot y) = p(y) \cdot p(x \subset y) = p(x) \cdot p(y \subset x).$$

We can divide both sides of this latter equation by $p(y)$ provided $p(y) \neq 0$. Hence

THEOREM 17. $p(x \subset y) = \dfrac{p(x) \cdot p(y \subset x)}{p(y)}$ *if* $x \neq 0$ *and* $p(y) \neq 0$.
($P$: 1. $T$: 16)

Theorem 17 is a form of Bayes' Theorem. To obtain a more usable form of this theorem we shall let $y$ be an element of $B$ such that $p(y) \neq 0$ and $x_1, x_2, \cdots, x_n$ be a set of mutually exclusive non-zero elements of $B$. We shall suppose that $y$ succeeds only if one of the events $x_1, x_2, \cdots, x_n$ succeeds, i. e., $y : \sim \cdot x_1 \vee x_2 \vee \cdots \vee x_n = 0$. From this equation it follows that

$$y = y \cdot x_1 : \vee : y \cdot x_2 : \vee : \cdots : \vee : y \cdot x_n$$

and hence

$$p(y) = p(y \cdot x_1) + p(y \cdot x_2) + \cdots + p(y \cdot x_n).$$

But by Theorem 16

$$p(y \cdot x_k) = p(x_k) \cdot p(y \subset x_k)$$

and hence

$$p(y) = \sum_{k=1}^{n} p(x_k) \cdot p(y \subset x_k).$$

Substituting this value of $p(y)$ in the equation of Theorem 17 and replacing $x$ by $x_i$ we obtain

THEOREM 18. *If $y$ is an element of $B$ such that $p(y) \neq 0$ and $x_1, x_2, \cdots, x_n$ are mutually exclusive non-zero elements of $B$ and*

$$y : \sim \cdot x_1 \vee x_2 \vee \cdots \vee x_n = 0,$$

*then*

$$p(x_i \subset y) = \frac{p(x_i) \cdot p(y \subset x_i)}{\sum_{k=1}^{n} p(x_k) \cdot p(y \subset x_k)}. \qquad (P: 1.\ T: 13, 16, 17)$$

Let $x_1, x_2, \cdots$ be the non-zero elements of $A$ (taken in order) such that $x \cdot x_k = x_k$. Then $x \cdot (n) = x_1 \vee x_2 \vee \cdots \vee x_m$ where

$$p[(m) \subset (n)] = p[x \subset (n)].$$

It follows from Theorem 7 and Postulate 14 that

$$p[y \subset : x \cdot (n)] = \sum_{k=1}^{m} p(y \subset x_k)/m$$
$$= p[y \subset x \cdot \subset (m)] = \sum_{k=1}^{m} p(y \subset x \cdot \subset k)/m.$$

Since the two summations are equal for all values of $n$ and hence for all values of $m$, it follows that

$$p(y \subset x_k) = p(y \subset x \cdot \subset k)$$

and therefore $y \subset x \cdot k = k$ if and only if $y \cdot x_k = x_k$. Thus

THEOREM 19. *If $x$ is an element of $B$ such that $x \neq 0$ and $x_1, x_2, \cdots$ are the non-zero elements of $A$ (taken in order) such that $x \cdot x_k = x_k$, then $y \subset x \cdot k = k$ if and only if $y \cdot x_k = x_k$.* $(P: 1, 3, 12, 14.\ T: 3, 7, 8)$

From Theorem 19 it follows that $\sim y \cdot \subset x : k = k$ if and only if $\sim y \cdot x_k = x_k$. But this latter equation holds if and only if $y \cdot x_k = 0$ and

hence if and only if $y \subset x \cdot k = 0$ and in turn if and only if $\sim \cdot y \subset x : k = k$. Therefore by Postulate 4

THEOREM 20. $\sim y \cdot \subset x = \sim \cdot y \subset x$ *if* $x \neq 0$. $(P: 1, 3, 4.\ T: 19)$

In a like manner the following two theorems are easily proved:

THEOREM 21. $x \cdot y : \subset z = x \subset z \cdot y \subset z$ *if* $z \neq 0$.
$(P: 1, 3, 4.\ T: 19)$

THEOREM 22. $x \vee y \cdot \subset z = x \subset z \cdot \vee \cdot y \subset z$ *if* $z \neq 0$.
$(P: 1, 3, 4.\ T: 19)$

Let $x$ and $y$ be two elements of $B$ such that $x \neq 0$ and $y \subset x = U$. Then $y \subset x \cdot k = k$ for all values of $k$ and hence by Theorem 19 $y \cdot x_k = x_k$ for all values of $k$. Therefore $x \cdot y = x$ and consequently

$$y \vee \sim x = y \cdot x \vee \sim x : \vee \sim x = x \vee : y \cdot \sim x : \vee \sim x = U.$$

Conversely if $y \vee \sim x = U$, we can multiply both sides of this equation by $x$ and obtain

$$x \cdot y : \vee : x \cdot \sim x = x \cdot y = U \cdot x = x.$$

Hence

$$y \subset x = y \cdot x : \subset x = x \subset x = U.$$

Therefore

THEOREM 23. *If $x \neq 0$, then $y \subset x = U$ if and only if $y \vee \sim x = U$.*
$(P: 1, 3, 4, 6.\ T: 19)$

Theorem 23 shows the relation between the symbol $\subset$ as defined by this postulate system and the Whitehead-Russell definition of implication. The equivalence breaks down when $x = 0$. That is, in our system a contradiction does not imply all propositions. In all other cases where implication is applied in symbolic logic it is equivalent to the operation $y \subset x$ defined in our system. It should be noted that the demand that a contradiction imply all propositions is not necessary to symbolic logic.

**3. Certain sub-sets of Boolean elements.** The following definitions will enable us to gain further insight into the nature of the system which we have developed.

DEFINITION 1. *Let $F$ consist of all finite disjunctions of elements of $A$ and the negatives of such disjunctions. Then the elements of $F$ are called finite elements.*

DEFINITION 2. *An element $x$ belongs to $P$ if and only if there exists a*

*non-zero atomic element* $\lambda$ *such that* $x \subset n = x \subset \cdot n + \lambda$ *for every non-zero atom* $n$. *The elements of* $P$ *are called periodic elements.*

DEFINITION 3. *An element* $u$ *belongs to* $R$ *if and only if there exist two elements* $x$ *and* $z$ *of* $F$ *and an element* $y$ *of* $P$ *such that* $u = x \cdot y : \vee z$. *The elements of* $R$ *are called rational elements.*

DEFINITION 4. *An element* $x$ *belongs to* $C$ *if and only if there exist atoms* $r$ *and* $n$ *such that* $x \cdot s = s$ *or* $0$ *according as* $s$ *is or is not of the form* $r + 1 + kn$ *where* $s$ *is an element of* $A$, $n > 0$ *and* $k = 0, 1, 2, \cdots$. *The element* $x$ *thus defined will be denoted by* $(r, n)$. *The elements of* $C$ *will be called fundamental rational elements.*

The elements of these various sets can most easily be described by means of their spectra. The spectrum of a finite element consists either of a finite number of 1's and an infinite number of 0's or of a finite number of 0's and an infinite number of 1's. The spectrum of a periodic element is periodic. The spectrum of a rational element may be irregular up to a certain point but from that point on it will be periodic. Thus the spectrum of a rational element is the sequence of digits of the binary development of some rational fraction. Conversely corresponding to any given rational fraction there is a rational element whose spectrum is the binary development of this fraction. The elements of $C$ are special rational elements. The spectrum of the element $(r, n)$ consists of $r$ 0's followed by a 1, then followed by $n - 1$ 0's next a 1, then $n - 1$ 0's, next a 1, etc.

It follows from Postulate 15 that if $x$ is any element of $B$, then $x \subset (n)$ is periodic with period $n$. The first $n$ terms of the spectra of $x$ and $x \subset (n)$ are the same. These $n$ terms are repeated periodically in the spectrum of $x \subset (n)$. Next let us specialize the element $x$. Let $0 \leqq r < n$. Then it is easily seen that $r + 1 \cdot \subset (n) = (r, n)$. We shall next investigate how the element $(r, n)$ can be constructed when $n \leqq r$. We shall consider first the element $(r, 1)$. In the spectrum of this element the first $r$ terms are 0's, the $(r + 1)$-st term is a 1, and every term thereafter is a 1. Hence $(r, 1) = \sim(r)$. In particular $(0, 1) = \sim(0) = \sim 0 = U$. It is now easily seen that if $0 \leqq r < n$, then $r + 1 \cdot \subset (n) : (kn, 1) = (kn + r, n)$. Hence any fundamental rational element $(r, n)$ can be constructed by means of the operations permitted within the system. An arbitrary rational element can be constructed by means of a finite disjunction of fundamental rational elements and atomic elements. Thus suppose we have the binary development of an arbitrary rational fraction. We can start with the atoms and by means of the operations permitted within the system construct a rational element whose spectrum is the given binary development.

The atoms and the rational elements may be considered as the analogues of the integers and the rational numbers respectively. Any finite combination of Boolean elements formed by means of the operators $\cdot$, $\vee$, $\sim$, $\subset$ in the manner permitted by the postulates is called a rational function of those elements. An arbitrary rational element can be constructed by forming the proper rational function of the proper atoms. Moreover any rational function of rational elements is a rational element. We have already observed that rational elements have ultimately periodic spectra analogous to the developments of rational numbers. The probability of a rational element is equal to the total number of successes in a period divided by the length of the period. For example if $x$ is a periodic element with period $\lambda$, then $p(x) = p[x \subset (\lambda)]$. The initial irregularity of a rational element does not effect its probability. Thus the probability of every rational element is rational. The following ten theorems are now easily proved.

THEOREM 24. *$R$ includes all the elements of $F$, $P$, and $C$.*

THEOREM 25. *If $0 \leqq r < n$, then $(r, n) = r + 1 \cdot \subset (n)$ and $(r, n)$ is an element of $P$.*

THEOREM 26. *$(r, 1) = \sim (r)$ and is an element of $F$.*

THEOREM 27. *$(0, 1) = U$.*

THEOREM 28. *If $0 \leqq r < n$, then $(kn + r, n) = r + 1 \cdot \subset (n) : (kn, 1)$ where $k = 0, 1, 2, \cdots$.*

THEOREM 29. *If $x$ is an element of $F$, then $x$ is a finite disjunction of elements of $A$ and $C$.*

THEOREM 30. *If $x$ is an element of $R$, then $x$ is a finite disjunction of elements of $A$ and $C$.*

THEOREM 31. *If $x$ is a non-zero element of $P$, then $x$ is a finite disjunction of elements of $C$.*

THEOREM 32. *$F$, $P$, and $R$ each satisfy Postulate 1 and hence constitute Boolean algebras.*

THEOREM 33. *Postulates 1-15 will be satisfied if $B$ is replaced by $R$.*

In previous papers (see Copeland I) the author has shown how the operations $\sim$, $\cdot$, $\vee$, $\subset$ can be interpreted as operations on spectra. The spectra can in turn be characterized by numbers. Thus in the light of Theorem 33 it is easy to construct an interpretation in which the elements are rational

numbers. This interpretation will satisfy all the postulates and hence the system is consistent.

**4. Relation to other systems.** The postulate system of this paper is obviously not categorical. This lack of completeness permits a certain flexibility which enables us to express other systems on the base of our postulates. However in order to do this it is necessary to include further definitions. We shall consider first the system of Kolmogoroff which is based on the concepts of field and distribution function. To this end we shall introduce the following definitions:

DEFINITION 5. *A system $B'$ is called a field*[4] *if it satisfies Postulate* 1.

DEFINITION 6. *A field $B'$ (with corresponding sub-set $A'$) is atomic if $B'$ and $A'$ satisfy Postulates* 2, 3, *and* 4.

DEFINITION 7. *A field $B'$ is a sub-field of $B$ if $B'$ is a sub-set of $B$.*

DEFINITION 8. *An atomic sub-field $B'$ of $B$ is complete provided the following condition is satisfied: if $x$ and $y$ are any two elements of $B$ such that $x \cdot z = y \cdot z$ for every element $z$ of $A'$, then $x = y$.*

Let $B'$ be a complete atomic sub-field of $B$ and let $x$ be any element of $A$. Then if $y$ is any element of $A'$, it will also be an element of $B$ and hence $x \cdot y = x$ or 0. But if $x \cdot y = 0$ for all elements $y$ of $A'$, then it follows from Definition 8 that $x = 0$. Thus if $x \neq 0$, there must exist an element $y$ of $A'$ such that $x \cdot y = x$. If on the other hand $x = 0$, then $x \cdot y = x$ for every $y$ of $A'$. Hence

THEOREM 34. *If $B'$ is a complete atomic sub-field of $B$ and if $x$ is any element of $A$, there exists an element $y$ of $A'$ such that $x \cdot y = x$.*

We shall now introduce the concept of distribution function.[5]

DEFINITION 9. *A function $\pi$ is a distribution function of a field $B'$ if*

(a) $\pi(x)$ *exists and is greater than or equal to* 0 *if $x$ is any element of $B'$.*

(b) $\pi(x \vee y) = \pi(x) + \pi(y) - \pi(x \cdot y)$ *for any two elements $x$ and $y$ of $B'$.*

(c) $\pi(U) = 1$ *and* $\pi(0) = 0$.

It will be observed that the set $B$ itself is a field and that the function $p$

---

[4] See Kolmogoroff I and Hausdorff I.

[5] See Kolmogoroff I.

is a distribution function of this field. Thus $B$ and $p$ satisfy Kolmogoroff's postulates. In Kolmogoroff's system conditional probabilities are provided for by the introduction of new fields whereas in our system $y \subset x$ is an element of $B$ if $x$ and $y$ are elements of $B$ and $x \neq 0$. However the field $B$ is somewhat restricted by Postulate 12. This restriction is to a certain extent alleviated in the application of the system by a device which permits us to establish a correspondence between a sub-field $B''$ of $B$ and a more general field $B'$. To this end we introduce the following definitions:

DEFINITION 10. *The following conditions are imposed on a correspondence $x \rightarrow y$ between the elements $x$ of a field $B''$ (with distribution function $p$) and the elements $y$ of a field $B'$ (with distribution function $\pi$). When such a correspondence exists, $B''$ is said to correspond to $B'$ and this fact is indicated by the symbol $B' \rightarrow B''$.*

(a) *If $x$ is an element of $B'$, there exists an element $y$ of $B''$ such that $x \rightarrow y$.*

(b) *If $y$ is an element of $B''$, there exists an element $x$ of $B'$ such that $x \rightarrow y$.*

(c) *If $x \rightarrow y$ and $x \rightarrow z$, then $y = z$.*

(d) *If $x \rightarrow y$, then $\sim x \rightarrow \sim y$.*

(e) *If $x \rightarrow u$ and $y \rightarrow v$, then $x \cdot y \rightarrow u \cdot v$.*

(f) *If $x \rightarrow y$, then $\pi(x) = p(y)$.*

DEFINITION 11. *The correspondence $B' \rightarrow B''$ is said to be atomic if $B'$ and $B''$ are atomic and if*

(g) *For every element $y$ of $A''$ there exists an element $x$ of $A'$ such that $x \rightarrow y$.*

Let $x$ and $y$ be elements of $B'$ and $u$ and $v$ be elements of $B''$ (where $B' \rightarrow B''$) and let $x \rightarrow u$ and $y \rightarrow v$. Then

$$x \vee y = \sim : \sim x \cdot \sim y \rightarrow \sim : \sim u \cdot \sim v = u \vee v.$$

Next let $x$ and $y$ be elements of $B'$ and let $z$ be an element of $B''$ such that $x \rightarrow z$ and $y \rightarrow z$. Then

$$x \cdot \sim y : \vee : y \cdot \sim x \rightarrow z \cdot \sim z : \vee : z \cdot \sim z = z \cdot \sim z.$$

Hence

$$\pi(x \cdot \sim y : \vee : y \cdot \sim x) = p(z \cdot \sim z) = 0.$$

We now have the following two theorems:

THEOREM 35. *If $x \rightarrow u$ and $y \rightarrow v$ where the correspondences satisfy the conditions of Definition* 10, *then $x \vee y \rightarrow u \vee v$.*

THEOREM 36. *If $x \rightarrow z$ and $y \rightarrow z$ where the correspondences satisfy the conditions of Definition* 10, *then $\pi(x \cdot \sim y : \vee : y \cdot \sim x) = 0$.*

The correspondence defined in Definition 10 is not one to one. However an element of $B'$ uniquely determines the corresponding element of $B''$. Furthermore if two elements of $B'$ correspond to the same element of $B''$, it follows from Theorem 36 that these two elements differ by an element whose probability is 0. Thus the correspondence preserves all probability relations and at the same time permits $B'$ to be a much more general field than $B''$. In order for the theory to include the mathematical counterpart of the observation it is necessary for the correspondence to be atomic. When this is the case, we can establish a correspondence between the elements of $A$ and the atomic elements of $B'$ in the following manner:

DEFINITION 12. *If $x$ is an element of $A$, $y$ an element of $A'$, $z$ an element of $A''$, $y \rightarrow z$, and $x \cdot z = x$, where $B''$ is a complete atomic sub-field of $B$ and the correspondence $B' \rightarrow B''$ is atomic; then $x \rightarrow y$.*

Let $x$ be an element of $A$ other than 0. Then there exists an element $z$ of $A''$ such that $x \cdot z = x$. The element $z$ satisfying this equation is unique since the elements of $A''$ are mutually exclusive. It follows from Definition 11 that there is an element $y$ of $A'$ such that $y \rightarrow z$, i. e., $x \rightarrow y$. Next let us suppose that there are two elements $y_1$ and $y_2$ of $A'$ such that $y_1 \rightarrow z$ and $y_2 \rightarrow z$. Then since the elements of $A'$ are mutually exclusive,

$$y_1 \cdot \sim y_1 = y_1 \cdot y_2 \rightarrow z \cdot z = z.$$

But $y_1 \cdot \sim y_1 \rightarrow z \cdot \sim z = 0$. Hence by Definition 10, $z = 0$. Since this is a contradiction, the correspondence $x \rightarrow y$ uniquely determines $y$. We now have the following theorems:

THEOREM 37. *For every element $x$ of $A$ there exists an element $y$ of $A'$ such that $x \rightarrow y$, where the sets $A$ and $A'$ and the correspondence $x \rightarrow y$ are defined as in Definition* 12.

THEOREM 38. *If $x \rightarrow y_1$, $x \rightarrow y_2$, and $x \neq 0$ where the correspondences are defined as in Definition* 12; *then $y_1 = y_2$.*

Let $u$, $v$, $x$, $y$, $z$ be respectively elements of $B'$, $B''$, $A$, $A'$, $A''$ such that $u \rightarrow v$, $x \rightarrow y$, $y \rightarrow z$, and $x \neq 0$. Then $u \cdot y \rightarrow v \cdot z$ and $x \cdot z = x$. Hence $v \cdot x = x$ if and only if $u \cdot y = y$. Thus

THEOREM 39. *If $u$, $v$, $x$, $y$ are respectively elements of $B'$, $B''$, $A$, $A'$ such that $u \rightarrow v$, $x \rightarrow y$, and $x \neq 0$ where $B'$, $B''$, $A$, $A'$ are related as in Definition* 12, *$u \rightarrow v$ is defined as in Definition* 10, *and $x \rightarrow y$ is defined as in Definition* 12; *then $v \cdot x = x$ if and only if $u \cdot y = y$.*

The set $A'$ can be interpreted as a space and the elements of $A'$ as points of this space. The set $B'$ is then a field of the space $A'$ and the elements of $B'$ are sets of points of the space. Definition 12 assigns to each non-zero element of $A$ a unique label (Merkmal[6]), i. e., a unique point of the space $A'$. The inverse correspondence is not unique. In fact to a point of $A'$ there may correspond an infinite number of elements of $A$. Moreover there may be points of $A'$ which correspond to no elements of $A$. Definition 12 merely demands that there be a sub-set $K$ of $A'$ such that to each element of $A$ there corresponds a unique element of $K$ and to each element of $K$ there correspond one or more elements of $A$. Since $A$ possesses a serial order, the correspondence determines a unique sequential arrangement of the elements of $K$. In the sequence $K$ the same element may be repeated a finite or an infinite number of times.

Let $u$ be an element of $B'$ and $v$ be an element of $B''$ such that $u \rightarrow v$. Of the first $n$ terms of the sequence $K$ let $m$ be the number which are elements of the set $u$. It then follows from Theorems 7 and 39 that $p[v \subset (n)] = m/n$. Then $m/n$ is the relative frequency with which an element of $u$ occurs in the first $n$ terms of $K$. We have $\lim_{n \rightarrow \infty} m/n = p(v) = \pi(u)$. Hence the sequence $K$ possesses the distribution (Verteilung) $\pi(u)$ defined over the field $B'$. The sequence $K$ can be interpreted as a spectrum of a much more general nature than hitherto considered. The terms of $K$ are the observations. These observations may be numbers, points, successes, failures, colors, or anything we please.

The set $B'$ is still somewhat less general than the type of field permitted by Kolmogoroff. It is nevertheless sufficiently general for the theory of probability. The extent to which $B'$ is restricted will be considered in detail in a subsequent paper. It will be sufficient here to note the results. We can start with an arbitrary denumerable field and extend it in the following manner. An element $x$ will belong to the extended field provided that for every positive number $\epsilon$ there exist two elements $y$ and $z$ of the denumerable field such that $x \cdot y = y$, $x \cdot z = x$, and $\pi(z \cdot \sim y) < \epsilon$.

We have observed that $K$ possesses one of the properties of a collective (Kollektiv[6]). Namely it has a distribution. The remaining properties of the collective are expressed in terms of the operation of selection (Auswahl[7]).

---

[6] See v. Mises I.

[7] v. Mises I.

We have seen that in the expreession $x \subset a$ we may regard $a$ as a selection operating on the element $x$. In order to study the operation of selection we shall introduce the following definitions:

DEFINITION 13. *If* $x \subset a \cdot \subset b = x \subset c$ *for every* $x$ *of* $B$, *then* $a \times b = c$ *and conversely.*

DEFINITION 14. *A set* $S$ *of elements of* $B$ *is called fundamental*[8] *if and only if*

(a) *$S$ contains $U$.*

(b) *If $a$ and $b$ are elements of $S$, then $a \times b$ is an element of $S$.*

The product $a \times b$ can most easily be studied by considering the corresponding spectra. In the spectrum of $A$ let $m_1, m_2, \cdots$ be the superscripts of the terms which are equal to 1 and let $n_1, n_2, \cdots$ be defined correspondingly for $b$. Then $m_{n_1}, m_{n_2}, \cdots$ indicate the positions of the 1's in the spectrum of $a \times b$. It will be noted that if the sequence $m_1, m_2, \cdots$ is finite and the sequence $n_1, n_2, \cdots$ is infinite, $a \times b$ is not defined. We shall, however, be interested in the cases where both sequences are infinite. In this case the product $a \times b$ always exists and is unique. Next if $\nu_1, \nu_2, \cdots$ indicate the positions of the 1's in the spectrum of $c$, then in the spectra of both $(a \times b) \times c$ and $a \times (b \times c)$, $m_{n\nu_1}, m_{n\nu_2}, \cdots$ are the superscripts of the terms which are equal to 1. Hence $(a \times b) \times c = a \times (b \times c)$. All the terms of the spectrum of $U$ are equal to 1 and therefore $a \times U = U \times a = a$. In the expression $x \subset : a \cdot b$ the terms of $x$ which are selected by the operator $a \cdot b$ are those which correspond to 1's in the spectra of both $a$ and $b$. This selection can be performed in an alternative manner. We first make the selections $x \subset b$ and $a \subset b$. If we now select from the spectrum of $x \subset b$ those terms which correspond to 1's in the spectrum of $a \subset b$, we thereby select from $x$ those terms which correspond to 1's in the spectra of both $a$ and $b$. Hence $x \subset : a \cdot b = x \subset b \cdot \subset \cdot a \subset b$. It follows from Definition 13 that $a \cdot b = b \times \cdot a \subset b = a \times \cdot b \subset a$. By considering the spectra of the elements $(r, n)$ and $(\rho, \nu)$ of the set $C$ (see Definition 4) it is easily proved that $(r, n) \times (\rho, \nu) = (r + \rho n, n\nu)$. We have already noted that $(0, 1) = U$. In the following theorems each of the elements $a, b, c$ is such that its spectrum contains an infinite number of 1's. This implies that these elements are different from 0.

THEOREM 40. $(a \times b) \times c = a \times (b \times c)$.

---

[8] Dörge I.

THEOREM 41. $a \times U = U \times a = a$.

THEOREM 42.

$$x \subset : a \cdot b = x \subset b \cdot \subset \cdot a \subset b \text{ and } a \cdot b = b \times \cdot a \subset b = a \times \cdot b \subset a.$$

THEOREM 43. *The set $C$ is fundamental,*

$$(0, 1) = U, \text{ and } (r, n) \times (\rho, \nu) = (r + \rho n, n\nu).$$

From Theorem 43 it follows that the product $a \times b$ is not in general commutative. Moreover the operation $a \times b$ does not in general have an inverse since when an element has once been rejected in the selection process, it can never be returned by further selection.

The following definitions enable us to study the systems of Wald, v. Mises, Reichenbach, and the author:

DEFINITION 15. *If $B_1, B_2, \cdots, B_n$ is a set of sub-fields of $B$, then these fields are independent provided $p(x_1 \cdot x_2 \cdots x_n) = p(x_1)p(x_2) \cdots p(x_n)$ for every set of elements $x_1, x_2, \cdots, x_n$ such that $x_i$ is an element of $B_i$ $(i = 1, 2, \cdots, n)$.*

DEFINITION 16. *If $B'$ is a sub-field of $B$ and $a$ is an element of $B$, then $B' \subset a$ is the field defined as follows:*

(a) *If $x$ is an element of $B'$, then $x \subset a$ is an element of $B' \subset a$.*

(b) *If $y$ is an element of $B' \subset a$, then there exists an element $x$ of $B'$ such that $x \subset a = y$.*

DEFINITION 17. *A sub-field $B'$ of $B$ is a collective field with respect to a fundamental set $S$ provided $p(x \subset y \cdot \subset a) = p(x \subset y)$ for every element $a$ of $S$ and every pair of elements $x$ and $y$ of $B'$.*

DEFINITION 18. *Let $B''$ be a collective field with respect to a fundamental set $S$ and let the correspondence $B' \to B''$ be atomic. Then since $A$ is an $\omega$-series, the correspondence defined by Definition 12 prescribes a sequential order for a sub-set $K$ of $A'$. The sequence $K$ thus defined is called a collective with respect to the fundamental set $S$.* (See Wald I.)

DEFINITION 19. *If the fundamental set $S$ (of Definition 18) satisfies the Regellosigkeitsprincip of v. Mises, then $K$ is called a collective.* (See v. Mises I.)

DEFINITION 20. *A sub-field $B'$ of $B$ is said to be admissible if*

(a) $p[x \subset (r, n)] = p(x)$ *for every element $x$ of $B'$ and for every pair of integers $r$ and $n$ such that $0 \leqq r < n$.*

(b) *The fields* $B' \subset (0, n)$, $B' \subset (1, n), \cdots, B' \subset (n-1, n)$ *are independent for every positive integer* $n$.

DEFINITION 21. *A sequence K defined in terms of an admissible field* $B''$ *in a manner similar to Definition* 18, *is called a normale Folge* (Reichenbach I) *or admissible sequence* (Copeland II).

DEFINITION 22. *If the field* $B''$ *satisfies condition Definition* 20 (a) *but not Definition* 20 (b), *then the corresponding sequence K is said to possess after-effect* (Reichenbach I).

DEFINITION 23. *A field* $B'$ *is called an admissible collective field with respect to a fundamental set S provided*

(a) *It satisfies Definition* 20.

(b) *Any operation of the form* $B' \subset x \cdot \subset a$ (*where* $x$ *and* $a$ *are respectively elements of* $B'$ *and* $S$) *transforms the admissible collective field* $B'$ *into a new admissible collective field* $B' \subset x \cdot \subset a$.

With the aid of the above definitions the systems of Kolmogoroff, Wald, v. Mises, Reichenbach, and the author can all be formulated on the base of our postulates. The system of Dörge could also be so expressed but a considerable number of additional definitions would be required.

UNIVERSITY OF MICHIGAN.

## REFERENCES.

B. A. Bernstein,

I. "Postulates for Boolean algebra involving the operation of complete disjunction," *Annals of Mathematics*, vol. 37 (1936), pp. 317-325.

A. H. Copeland,

I. "Predictions and probabilities," *Erkenntnis*, vol. 6, pp. 189-203.

II. "Admissible numbers in the theory of geometrical probability," *American Journal of Mathematics*, vol. 53 (1931), pp. 153-162.

K. Dörge,

I. "Zu der von R. v. Mises gegebenen Begründung der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift*, vol. 40 (1936), pp. 161-193.

H. P. Evans and S. C. Kleene,

I. "A postulational basis for probability," *American Mathematical Monthly*, vol. 46 (1939), pp. 141-148.

F. Hausdorff,

I. *Mengenlehre,* 2nd ed., Berlin-Leipzig (1927).

E. V. Huntington,

I. "Sets of independent postulates for the algebra of logic," *Transactions of the American Mathematical Society,* vol. 5 (1904), p. 309.

II. *The Fundamental Propositions of Algebra.* Monograph 4 of J. W. A. Young's Monographs on Modern Mathematics, New York, 1911.

A. Kolmogoroff,

I. "Grundbegriffe der Wahrscheinlichkeitsrechnung," *Ergebnisse der Mathematik,* vol. 2, no. 3.

B. O. Koopman,

I. "The axioms of intuitive probability," *Annals of Mathematics,* vol. 41 (1940), pp. 259-292.

II. "The bases of probability," *Bulletin of the American Mathematical Society,* vol. 46 (1940), pp. 763-774.

III. "Intuitive probabilities and sequences," *Annals of Mathematics,* vol. 42 (1941), pp. 169-187.

R. von Mises,

I. "Grundlagen der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift,* vol. 5 (1919), pp. 52-99.

H. Reichenbach,

I. "Axiomatik der Wahrscheinlichkeitsrechnung," *Mathematische Zeitschrift,* vol. 34 (1932), pp. 568-619.

Whitehead and Russell,

I. *Principia Mathematica,* vol. 1, Cambridge University Press (1925).

# ON THE NUMBER OF REPRESENTATIONS OF THE SQUARE OF AN INTEGER AS THE SUM OF THREE SQUARES.*

By C. D. OLDS.

**1. Introduction.** Let $N_r(n)$ denote the number of representations of an integer $n$ as the sum of $r$ squares. Two representations

$$n = x_1^2 + x_2^2 + \cdots + x_r^2, \qquad n = y_1^2 + y_2^2 + \cdots + y_r^2,$$

are distinct unless simultaneously $x_\nu = y_\nu$, $\nu = 1, 2, \cdots, r$. Notice that in a given representation the signs of the roots as well as their arrangement are relevant. A zero square is considered positive.

In 1884, T. J. Stieltjes [7] [1] proved by means of elliptic functions that $N_3(p^{2a}) = 6p^a$, where $p$ is a prime of the form $8k+1$. Later, in 1907, A. Hurwitz [2] stated without giving the proof, that if [2]

$$(1.1) \qquad n^2 = 2^{2a}m = 2^{2a}P^2Q^2, \quad m = P^2Q^2 \equiv 1 \pmod 8, \quad Q = \prod_{\nu=1}^{s} q_\nu^{a_\nu},$$

where $q_1, q_2, \cdots, q_s$ are distinct primes of the form $4k+3$, and where $P$ is the product of primes of the form $4k+1$, then

$$(1.2) \qquad N_3(n^2) = 6P \prod_{\nu=1}^{s} [q_\nu, a_\nu],$$

where

$$[q_\nu, a_\nu] = q_\nu^{a_\nu} + 2(q_\nu^{a_\nu} - 1)/(q_\nu - 1).$$

This result was also stated by Stieltjes [7]. G. Pall [5] gave an analytic proof of (1.2) in 1930, and an arithmetical proof [6] using quaternions in 1940.

The author [4] has given an arithmetical derivation of (1.2) using a method similar to that used by Hurwitz [3] in his proof of the analogous formula for $N_5(n^2)$. In this paper the same end is accomplished by using a method which may be regarded as a generalization of that used by Stieltjes [7] in his discussion of the special case $N_3(p^{2a})$.

**2. Three lemmas.** The work which follows is based upon an identity which has been proved arithmetically by J. V. Uspensky [8]. We call it

* Received March 4, 1941; this is the third part of a paper presented to the American Mathematical Society April 6, 1940, under the title "On the number of representations of the square of an integer as the sum of an odd number of squares." The author wishes to thank Professor J. V. Uspensky for help in preparing this paper.

[1] The numbers in brackets refer to the bibliography.

[2] The meaning of $n$, $m$, $P$, and $Q$ given in (1.1) will hold throughout the paper.

LEMMA 2.1. *Let $F(x,y,z)$ be an arbitrary function defined for integral values of the arguments, and satisfying the following conditions of parity:*

$$F(-x,y,z)=-F(x,y,z),\quad F(x,-y,-z)=F(x,y,z),\quad F(0,y,z)=0.$$

*Then for any positive integer $m\equiv 1 \pmod 8$,*

$$\sum_{(a)} F(\Delta'+2\Delta,h,\Delta'-2\Delta)=\sum_{(b)} F(i+2d,i-2\delta,i+2d-2\delta)+T \tag{2.1}$$

*where*

$$T=\{\sum_{(c)}[F(s,j,j)-F(j,s,j)]\},\qquad m=s^2,\ s>0.$$

*The first two sums extend respectively over integral solutions of the equations*

$$\begin{array}{ll}\text{(a)} & m=h^2+8\Delta\Delta',\ \Delta' \textit{ odd};\ \Delta,\Delta' \textit{ both positive};\\ \text{(b)} & m=i^2+8d\delta,\ \delta \textit{ odd};\ d,\delta \textit{ both positive};\end{array}$$

*and the third summation runs through values*

$$\text{(c)}\quad j=1,3,5,\cdots,s-2,\ m=s^2,\ s>0.$$

*In $T$ the symbolic notation $\{A\}$ is equal to $A$ itself when $m$ is a perfect square: $m=s^2$, $s>0$, and is zero otherwise.*

From the equation (2.1), by placing suitable restrictions upon the highly arbitrary function $F(x,y,z)$, many interesting relations may be derived. We are interested in the one which follows. In (2.1) take

$$F(x,y,z)=(-1)^{(x+y-z-1)/2}\phi(x,y),$$

where $\phi(x,y)$ is an arbitrary function satisfying the conditions

$$\phi(x,y)=\phi(-x,y)=\phi(x,-y).$$

The result of this substitution is

$$\sum_{(b)}(-1)^{(i-1)/2}\phi(i+2d,i-2\delta) = \{\sum_{(c)}(-1)^{(s-1)/2}[\phi(j,s)-\phi(s,j)]\}. \tag{2.2}$$

If we now set

$$\phi(x,y)=\cos(\pi y/4),$$

we get in place of (2.2)

$$2\sum_{(b)}(-1)^{(i^2-1)/8+(\delta-1)/2} = \{(s-1)(-1)^{(s-1)/2+(s^2-1)/8}-2(-1)^{(s-1)/2}\sum_{(c)}(-1)^{(j^2-1)/8}\},$$

which is easily reduced to [3]

LEMMA 2.2. $4\sum_{(d)}(2/i)\rho(m-i^2)=G(m),\qquad m-i^2\geqq 0,$

---

[3] The symbol $(P/Q)$ in Lemma 2.2 is the Jacobi symbol.

*where* $(d)$ *indicates that* $i$ *runs over all positive odd integers rendering* $m - i^2 \geqq 0$, *and where*

$$G(m) = (-2/s)s \text{ if } m = s^2,\ s > 0;\quad G(m) = 0 \text{ otherwise.}$$

*The function* $\rho(n)$ *is defined as follows:*

$$\rho(n) = \sum_{n=d\delta,\ \delta \text{ odd}} (-1)^{(\delta-1)/2}, \qquad \rho(0) = 1/4,$$

*the summation ranging over all odd integers* $\delta$ *satisfying the equation* $n = d\delta$.

For convenience set

$$\Phi(Q) = 2\sum_{(d)} \rho(m - i^2), \qquad \Psi(Q) = 4\sum_{(d)} (2/i)\rho(m - i^2),$$

$$\phi(Q) = 2\sum_{(d)}{}^{*} \rho(m - i^2), \qquad \psi(Q) = 4\sum_{(d)}{}^{*} (2/i)\rho(m - i^2).$$

In the last two expressions the asterisk indicates that the summations extend over only those values of $i$ which are relatively prime to $Q$. Interesting relations exist between the functions just defined. Let $d = (i, Q)$ be the greatest common divisor of $i$ and $Q$, and write $i = de$ and $R = Q/d$, then $(e, R) = 1$. As $d$ runs through the divisors of $Q$, $R$ runs through the same set of divisors, but in a different order; hence we can write, recalling that $m = P^2Q^2$,

$$\Phi(Q) = 2\sum_{d|Q} \sum_{e=1,3,5,\ldots} \rho(P^2R^2d^2 - e^2R^2), \qquad (e, R) = 1$$

$$= \sum_{d|Q} [2\sum_{(d)}{}^{*} \rho(P^2d^2 - i^2)] = \sum_{d|Q} \phi(d),$$

where in order to obtain the last line we have to use the fact that $\rho(a^2b) = \rho(b)$ provided $a$ is only divisible by primes of the form $4k + 3$. We have found that

$$\Phi(Q) = \sum_{d|Q} \phi(d),$$

whence, by inversion,

$$\phi(Q) = \sum_{d|Q} \mu(d)\Phi(Q/d). \tag{2.3}$$

Here $\mu(d)$ is the well-known Möbius function. The analogous relationship between $\Psi(Q)$ and $\psi(Q)$ is obtained in exactly the same way. It is

$$(2/Q)\Psi(Q) = \sum_{d|Q} (2/d)\psi(d).$$

This leads, by inversion, to

$$\psi(Q) = \sum_{d|Q} \mu(d)(2/d)\Psi(Q/d). \tag{2.4}$$

The next step is to evaluate $\phi(Q) \pm \psi(Q)$ in two ways, first by using their defining equations, and second by combining (2.3) with (2.4). Equating the two expressions thus obtained, and reducing, we get

$$\sum_{d|Q} \mu(d)\Phi(Q/d) + \frac{\epsilon}{2}\sum_{d|Q} \mu(d)(2/d)\Psi(Q/d) = 4\sum{}^{*}\rho(m - i^2), \tag{2.5}$$

where we take $\epsilon = +1$ or $-1$ according as the sum on the right runs through values of $i \equiv \pm 1 \pmod 8$ or $i \equiv \pm 3 \pmod 8$. To transform (2.5) further we have two cases to consider according as $PQ \equiv 1, 3 \pmod 8$ or $PQ \equiv 5, 7 \pmod 8$. When $PQ \equiv 1$ or $3 \pmod 8$ and $i \equiv \pm 3 \pmod 8$, or when $PQ \equiv 5, 7 \pmod 8$ and $i \equiv \pm 1 \pmod 8$, then in either case one of the numbers $PQ - i$ or $PQ + i$ will be of the form $8k + 6$ and hence will contain an odd number of prime factors (not necessarily distinct) which are of the form $4k + 3$. Consequently, in either case, $m - i^2$ will contain primes of the form $4k + 3$ in odd power, and because of this $\rho(m - i^2)$ will be zero. It is clear, therefore, that we can replace (2.5) by

$$\sum_{d|Q} \mu(d)\Phi(Q/d) = \tfrac{1}{2}(-2/PQ) \sum_{d|Q} \mu(d)(2/d)\Psi(Q/d),$$

since, we recall, $(-2/PQ) = +1$ or $-1$ according as $PQ \equiv 1, 3 \pmod 8$ or $PQ \equiv 5, 7 \pmod 8$. From this last equation it follows that [4]

$$\sum_{d|Q} \mu(d)\Phi(Q/d) = (P/2) \sum_{d|Q} \mu(d)(-1/d)(Q/d) = (P/2)\Omega(Q),$$

whence, by inversion,

$$\Phi(Q) = (P/2) \sum_{d|Q} \Omega(d).$$

But, on the other hand,

$$\Omega(Q) = \Omega(q_1^{a_1} q_2^{a_2} \cdots q_s^{a_s}) = \prod_{\nu=1}^{s} \Omega(q_\nu^{a_\nu}),$$

where

$$\Omega(q_\nu^{a_\nu}) = \sum_{d|q_\nu^{a_\nu}} \mu(d)(-1/d)(q_\nu^{a_\nu}/d) = q_\nu^{a_\nu} + q_\nu^{a_\nu - 1} = f(q_\nu^{a_\nu}).$$

Thus

$$\Phi(Q) = (P/2) \sum_{d|Q} \Omega(d) = (P/2) \prod_{\nu=1}^{s} [1 + f(q_\nu) + \cdots + f(q_\nu^{a_\nu})]$$

$$= (P/2) \prod_{\nu=1}^{s} [q_\nu^{a_\nu} + 2(q_\nu^{a_\nu} - 1)/(q_\nu - 1)].$$

We have established

LEMMA 2.3. $\quad \Phi(Q) = (P/2) \prod_{\nu=1}^{s} [p_\nu, a_\nu].$

**3. The formula for $N_3(n^2)$.** To begin with, we see that

$$N_3(n^2) = N_3(2^{2a} m) = N_3(m).$$

In order to evaluate $N_3(m)$, notice that, by hypothesis, $m \equiv 1 \pmod 8$, and consequently among the roots of the equation

$$m = i^2 + j^2 + k^2,$$

---

[4] From now on the symbols $(P/2)$, $(Q/d)$ and $(q_\nu^{a_\nu}/d)$ denote fractions and not the Jacobi symbol.

one root, say $i$, must be odd, while the other two must be even. If $S$ represents the number of solutions of this equation in which $i$ is odd, then it is an easy matter to verify that

$$S = 2\sum N_2(m - i^2), \qquad (i = 1, 3, 5, \cdots),$$

the summation extending over all odd integers rendering the argument non-negative. It follows, therefore, that

$$N_3(m) = 3S = 6\sum_{(d)} N_2(m - i^2).$$

On the other hand it is well known that [5]

$$N_2(m) = 4\rho(m), \qquad \rho(0) = 1/4,$$

so that

$$N_3(n^2) = N_3(m) = 24\sum_{(d)} \rho(m - i^2) = 12\Phi(Q).$$

Now apply Lemma 2. 3. We get

$$N_3(n^2) = N_3(2^{2a}P^2Q^2) = 6P\prod_{\nu=1}^{s} [q_\nu, a_\nu],$$

which is the desired result (1. 2).

PURDUE UNIVERSITY,
LAFAYETTE, INDIANA.

## BIBLIOGRAPHY

1. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford, 1938, pp. 241-242.
2. A. Hurwitz, *L'Intermédiaire des Mathématiques*, vol. 14 (1907), p. 107.
3. ———, *Mathematische Werke*, vol. 2, Basel, 1933, pp. 5-7.
4. C. D. Olds, "On the representations, $N_3(n^2)$," *Bulletin of the American Mathematical Society*, vol. 47 (1941), pp. 499-503.
5. G. Pall, "On the number of representations of a square, or a constant times a square, as the sum of an odd number of squares," *Journal of the London Mathematical Society*, vol. 5 (1930), pp. 102-105.
6. ———, "On the arithmetic of quaternions," *Transactions of the American Mathematical Society*, vol. 47 (1940), pp. 487-500.
7. T. J. Stieltjes, *Correspondance d'Hermite et de Stieltjes*, Paris, 1905, vol. 1, letter 45, pp. 89-94.
8. J. V. Uspensky, "Sur les relations entre les nombres des classes des formes quadratiques binaires et positives," *Bulletin de l'Académie des Sciences de l'URSS*, Third memoire (1925), p. 328.

[5] For an arithmetical proof see Hardy and Wright [1].

# ON 1-BOUNDING MONOTONIC TRANSFORMATIONS WHICH ARE EQUIVALENT TO HOMEOMORPHISMS.* [1]

By EDITH R. SCHNECKENBURGER.

**1. Introduction.** A monotonic [2] transformation of a space $A$ on to a space $B$ is a transformation such that the inverse of each point $b$ of $B$ is connected; in other words, the 0-dimensional Betti number of the inverse of $b$ is zero. In this paper such a transformation will frequently be called an $R_0{}^0$-transformation. It could also be called a 0-bounding transformation.

A 1-bounding transformation, or $R_0{}^1$-transformation, of a space $A$ on to a space $B$ is a transformation such that for each point $b$ of $B$ the 1-dimensional Betti number of the inverse of $b$ is zero.

An $R_0{}^{01}$-transformation is a transformation which is monotonic and 1-bounding. In general, an $R_0{}^{01\cdots n}$-transformation of a space $A$ on to a space $B$ is a transformation such that for every point $b$ of $B$ the $i$-dimensional Betti number, $i = 0, \cdots, n$, of the inverse of $b$ is zero.

In this paper we consider a single-valued, continuous, $R_0{}^{01}$-transformation $T$ of a space $A$ on to a space $B$ such that $T$ is equivalent to a homeomorphism. We shall say that a transformation $T(A) = B$ is equivalent to a homeomorphism when $B$ is homeomorphic to $A$.

In a paper [3] by Roberts and Steenrod it is shown that a single-valued, continuous, $R_0{}^{01}$-transformation of $M$, a compact two-manifold without boundary, gives a space homeomorphic to $M$. This is a generalization of a theorem of R. L. Moore which states this result for the case in which $M$ is a two-sphere.[4] These results suggest the consideration of the following questions:

1. What spaces, other than two-manifolds, satisfy the condition that every $T$ of the type under consideration is equivalent to a homeomorphism?

2. What conditions must be placed upon a particular $T$ to make it equivalent to a homeomorphism on a given space $A$?

In section 3 we state conclusions of the nature described in question 2. In considering question 1, we have confined our attention to the study of a

* Received July 29, 1940; revised March 4, 1941.

[1] Presented to the American Mathematical Society, September 7, 1939.

[2] Defined as a monotone transformation by C. B. Morrey, *American Journal of Mathematics*, vol. 57 (1935), p. 18.

[3] J. H. Roberts and N. E. Steenrod, *Annals of Mathematics*, vol. 39 (1938), Theorem 1, p. 854.

[4] R. L. Moore, *Transactions of the American Mathematical Society*, vol. 27 (1925), Theorem 25, p. 247.

space $A$ which is a Peanian continuum imbedded in a two-manifold $M$. Our results are:

(a) If $A$ is cyclic, $A$ is a simple closed curve or the two-manifold $M$.

(b) If $A$ is acyclic, $A$ is an arc. (This is true for any acyclic Peanian continuum.)

(c) If $A$ is neither cyclic nor acyclic and certain restrictions are put upon the cut points of $A$, then $A$ is a finite or countably infinite set of simple closed curves tangent to each other at a single point $p$.

We have concluded that a complete characterization of $A$, if possible, would be too lengthy and tedious to be interesting.

The author wishes to thank Professor W. L. Ayres of the University of Michigan who suggested the investigation of this problem and gave many helpful criticisms and suggestions during the preparation of this paper.

**2. Notation.** In this paper, the following notation is used: $H \equiv$ homeomorphism; $T \equiv$ single-valued, continuous, $R_0^{01}$-transformation; $X \equiv$ Peanian continuum contained in a compact metric space; $M \equiv$ compact two-manifold without boundary; $A \equiv X$ imbedded in $M$; $B \equiv T(X)$ where $B$ is contained in a metric space, and it is assumed that $B$ is never a single point.

We shall also have occasion to use these symbols: $R^i(S) \equiv$ mod 2, $i$-dimensional Betti number of a set $S$; $\langle xy \rangle \equiv xy - x - y$ where $xy$ is an arc with endpoints $x$ and $y$; $xy\rangle \equiv$ arc $xy$ minus end point $y$; $\langle xy \equiv$ arc $xy$ minus end point $x$.

A knowledge of the cyclic element theory, as developed in a paper by Kuratowski and Whyburn,[5] will be presupposed.

**3. Conditions that a particular transformation be equivalent to a homeomorphism.** A vertex $y$ of a finite graph $X$ will be called *non-essential* if it is the end point of exactly two 1-cells $xy$ and $yz$ and $x \neq z$; otherwise, $y$ is an essential vertex.

LEMMA 1. *The 1-cells of any finite graph $X$ may be so re-defined that every vertex is essential.*

THEOREM 1. *If $X$ is a finite connected graph, a necessary and sufficient condition that $B = T(X)$ be homeomorphic to $X$ is that $T(u) \neq T(y)$ for any two essential vertices $u$ and $y$ of order $k \neq 2$.*

COROLLARY. *If $X$ is a simple closed curve, $B = T(X)$ is also a simple closed curve.*[6]

---

[5] C. Kuratowski and G. T. Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331.

[6] G. T. Whyburn, *Bulletin of the American Mathematical Society*, vol. 42 (1936), p. 66.

LEMMA 2. *If $H_0$ is an order preserving homeomorphism between a closed subset $S$ of an arc $\alpha = a_1a_2$ and a subset $H_0(S)$ of an arc $\beta = b_1b_2$ and $a_1 + a_2 \subset S$, $b_1 + b_2 \subset H_0(S)$, then there exists a homeomorphism $H$ between $\alpha$ and $\beta$ which agrees with $H_0$ on $S$.*

THEOREM 2. *If $X$ is a tree, $S$ is the set of all end points and branch points of $X$ and $T$ is a single-valued, continuous, $R_0{}^0$-transformation such that $T(u) \neq T(y)$ where $u$ and $y$ are any two points of $\bar{S}$, then $B = T(X)$ is a tree which is homeomorphic to $X$.*

The proofs of the lemmas and theorems, stated above, are not included in this paper since the method of proof in each case is quite obvious. Lemma 1 is used in the proof of Theorem 1 and Lemma 2 in the proof of Theorem 2.

**4. Conditions that every transformation be equivalent to a homeomorphism.**

LEMMA 3. *If $A$ is a proper subset of $M$ which is cyclicly connected but is not a simple closed curve, there is an acyclic continuum in $A$ which separates $A$.*

Let $x$ be any point of $A$ which is on the boundary of some complementary domain of $A$ in $M$. Choose an open 2-cell neighborhood $E$ of $x$ in $M$ such that: (1) $A \cdot (M - \bar{E}) \neq 0$, and (2) $A \cdot (M - \bar{E}) \neq \langle xy \rangle$, where $x + y \subset \bar{E}$. Since $A$ is not a simple closed curve, it is always possible to select an $E$ satisfying condition (2). Denote by $C$ the simple closed curve which is the boundary of $E$ and by $N$ the component of $A - A \cdot C$ which contains $x$. Then $\bar{N}$ is a continuous curve[7] which is contained in $E + C$.

It follows from a theorem of W. L. Ayres[8] that a point $q$ of $C \cdot \bar{N}$ is separated from $x$ in $\bar{N}$ by a finite number $n$ of arcs[9] $\alpha_i$ of $\bar{N}$, where $n$ is the number of components of $\bar{N} - x - q$ which have $x$ and $q$ as limit points. Let $N_q$ be the component of $\bar{N} - \sum_{i=1}^{n} \alpha_i$ which contains $q$. The set of components $N_q$ for all points of $C \cdot \bar{N}$ is a collection of open sets covering the closed and compact set $C \cdot \bar{N}$. By the Heine-Borel theorem a finite number of these components cover $C \cdot \bar{N}$. Denote these components by $N_i$.

For each $N_i$ there is a finite number of arcs $\alpha_{ji}$ such that $\sum_j \alpha_{ji} = W_i$ separates $x$ from $C \cdot N_i$ in $\bar{N}$. If, for some $i \neq k$, $\alpha_{ji} \cdot \alpha_{hk} \neq 0$ and the set

[7] R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, theorem 57, p. 232.

[8] W. L. Ayres, *Proceedings of the National Academy of Sciences*, vol. 14 (1928), theorem 2, p. 205.

[9] An arc may be a single point.

$\alpha_{ji} + \alpha_{hk} = V$ is not acyclic, we replace $V$ by a subcontinuum $V'$ which is acyclic. The set $V'$ is obtained by removing from each simple closed curve of $V$ the interior points of a subarc which contains no branch points. There is such an arc in each simple closed curve of $V$ because the number of arcs $\alpha_{hk}$ is finite and hence there can be at most a denumerable number of branch points of $V$ on any simple closed curve contained in $V$. It is evident that $V'$ is connected, closed, and acyclic.

We repeat the above process for every pair of arcs of $W = \sum_i W_i$ whose sum contains a simple closed curve. Thus, we obtain a finite set of acyclic continua $\beta_i$, $i = 1, \cdots, s$, $\beta_i \cdot \beta_j = 0$, $i \neq j$, where $\beta_i$ is either an arc $\alpha_{hk}$ or an acyclic continuum contained in the sum of two or more of these arcs.

Then $\bar{N} - \sum_{i=1}^{s} \beta_i = R + Q$, $\bar{R} \cdot Q = R \cdot \bar{Q} = 0$, $R$ connected, $x \in R$ and $C \cdot \bar{N} \subset Q$. Let $x_1 b_1$ be an arc of $R + \sum_{i=1}^{s} \beta_i$ such that $x_1 b_1\rangle \subset R$, $b_1 \in \beta_1$. In general, denote by $x_i b_i$, $i = 2, \cdots, s$, an arc such that $\langle x_i b_i\rangle \subset R - \sum_{j=1}^{i-1} x_j b_j$, $x_i$ a point of $\sum_{j=1}^{i-1} x_j b_j$, and $b_i$ a point of $\beta_i$. Let $A_1 = \sum_{i=1}^{s} (\beta_i + x_i b_i)$. The set $A_1$ is closed because it consists of a finite number of acyclic continua. Since each $x_i b_i$ has one and only one point in $\sum_{j=1}^{i-1} (\beta_j + x_j b_j)$, $A_1$ is connected and acyclic.

There are two cases to be considered. Let us suppose first that $R + \sum_{i=1}^{s} \beta_i \not\subset A_1$. Then, $\bar{N} - A_1 = S_1 + S_2$, $\bar{S}_1 \cdot S_2 = S_1 \cdot \bar{S}_2 = 0$, $R - A_1 \cdot R \subset S_1$, $C \cdot \bar{N} \subset S_2$. Since any arc joining $R - A_1 \cdot R$ to $A \cdot (M - \bar{E})$ must contain a point of $C \cdot \bar{N}$, the set $A_1$ is an acyclic continuum separating $A$.

In the second case, $A_1 = R + \sum_{i=1}^{s} \beta_i$. Let $z$ be a point of order 2 of $R$, $z_1 z z_2$ any subarc of $R$ containing no branch point of $A_1$, and $A'_1$ and $A''_1$ the components of $A_1 - z_1 z z_2$. Since $A$ is cyclicly connected, there is an arc $\rho$ which joins $A'_1$ and $A''_1$ and does not contain $z$. This arc $\rho$ contains a subarc $\rho_1$ having only its endpoints in $A'_1$ and $A''_1$. These endpoints must be contained in $\sum_{i=1}^{s} \beta_i$, hence, $\rho_1 \cdot (z_1 z_2) = 0$. Let $A_2 = A_1 - \langle z_1 z_2 \rangle + \rho_1$. Then $A - A_2 - z_1 z_2 \neq 0$ because it follows from condition (2) that $A - A \cdot \bar{E}$ cannot be a subarc of $\rho_1$. Hence, the acyclic continuum $A_2$ separates $\langle z_1 z_2 \rangle$ from points of $A - A_2 - z_1 z_2$. Thus, in either case, there is an acyclic continuum in $A$ which separates $A$.

THEOREM 3. *If $A$ is cyclicly connected and $B = T(A)$ is homeomorphic to $A$ for every $T$, either $A \equiv M$ or $A$ is a simple closed curve.*

The fact that $A$ can be the two-manifold $M$ follows from a theorem of Roberts and Steenrod.[3] Hence, we need consider only the case in which $A$ is a proper subset of $M$.

We suppose that $A$ is not a simple closed curve. By Lemma 3 there is an acyclic continuum $A_1$ in $A$ which separates $A$. Define $T(A)$ as follows: $T(A_1)$ is a single point, $T$ is a homeomorphism on $A - A_1$. Then $T(A_1)$ is a cut point of $B$ and $B$ is not homeomorphic to $A$ because $A$ is cyclicly connected.

Hence, if there exists any set $A$ which is a proper subset of $M$ and satisfies the hypothesis of the theorem, $A$ must be a simple closed curve. The existence of $A$ follows from the Corollary to Theorem 1.

THEOREM 4. *If $X$ is acyclic and $B = T(X)$ is homeomorphic to $X$ for every $T$, then $X$ is an arc.*

Let $\alpha_1$ be any arc of $X$ joining two end points of $X$, $\sum_i r_i$ the set of branch points of $X$ which are contained in $\alpha_1$, and $S_i$ the sum of all components $D_{ij}$ of $X - r_i$ such that $D_{ij} \cdot \alpha_1 = 0$. Define $T(X)$ as follows: $T$ is a homeomorphism on $\alpha_1$, $T(S_i) = T(r_i)$. Then $T$ is a single-valued, continuous, $R_0^{01}$-transformation. Further, since $T(X) = T(\alpha_1) = B$ and $T$ is a homeomorphism on $\alpha_1$, $B$ is an arc. Then $B$ cannot be homeomorphic to $X$ unless $X$ is also an arc. Thus $S_i = 0$ for every $i$. Since every transform of an arc is an arc when $T$ is monotonic,[6] $B$ is homeomorphic to $X$ for every $T$ when $X$ is an arc. Hence, if $X$ is acyclic, it is an arc.

THEOREM 5. *If $B = T(X)$ is homeomorphic to $X$ for every $T$, then the closed extension of every component of $X - C_k$ contains at least one $C_i$, where $C_j$ is a true cyclic element of $X$.*

Let $x_j$ be any cut point of $X$ which is contained in some $C_k$ and let $D_{jh}$ be any component of $X - C_k$ which has $x_j$ as limit point. Denote by $D_j$ the sum of all the components $D_{jh}$ such that $\bar{D}_{jh}$ contains no $C_i$. We shall prove that $D_j$ is vacuous. Define a transformation $T(X) = B$ as follows: $T(D_j + x_j) = x_j$, $T(x) = x$ where $x$ is any point of $X - \sum_j D_j$.

It can be shown that $T$ is a single-valued, continuous, $R_0^{01}$-transformation. Obviously, $T$ is single-valued. It is also evident that $T$ is continuous over any set of points contained in one set $D_j$ and over $X - \sum_j D_j$. Hence, to prove that $T$ is continuous it is necessary to consider only sequences of points $\{d_j\}$, where $d_j$ is a point of $D_j$. Let $d$ be the sequential limit point of such a sequence $d_{j_i}$. Then $d$ is contained in $X - \sum_j D_j$ because each set $D_k$ is open and can contain no limit point of $X - D_j$. Therefore, $T(d) = d$. Since for every $\epsilon > 0$ there is a $\delta > 0$ such that $S(d, \delta)$ is contained in a connected

subset of $S(d, \epsilon)$ and since the cut point $x_{j_i}$ separates $d$ and $d_{j_i}$, it follows that all but a finite number of $x_{j_i}$ are contained in $S(d, \epsilon)$; that is, $d$ is a sequential limit point of $\{x_{j_i}\}$. Hence, $T(d) = d$ is a limit point of $\{T(x_{j_i})\} = \{x_{j_i}\} = \{T(d_{j_i})\}$. This proves that $T$ is continuous.

The continuum $D_j + x_j$ contains only cyclic elements, namely, cut points and end points of $X$, and is therefore a continuous curve.[10] From this fact and the fact that $D_j + x_j$ is defined so that it contains no simple closed curve, it follows that $R^1(D_j + x_j) = 0$. Thus, $T$ satisfies all the required conditions.

Since $T$ is the identity transformation on each $C_i$, $T(C_i)$ is a true cyclic element of $B$. Furthermore, $B_i$, any true cyclic element of $B$, is contained in $T(C_j)$ for some $j$.[11] Hence, $B_i \equiv T(C_j)$. Let $b_k$ be any cut point of $B$ which is contained in some $B_k$. The closed extension of every component of $B - b_k$ contains at least one $T(C_j) = B_i$, because $T$ is homeomorphic on $X - \sum_j D_j$. Thus, $B$ is not homeomorphic to $X$ unless $D_{jh}$ of $X$ for every $j$ and $h$ satisfies the condition that $\bar{D}_{jh}$ contains at least one $C_i$.

LEMMA 4. *If $S$ is a closed subset of $X$ and each component $S_i$ of $S$ is contained in a continuum $L_i$ of $X$ where $R^1(L_i) = 0$ and $L_i \cdot L_j = 0$, $i \neq j$, then $S$ is contained in a continuum $W$ of $X$ such that $R^1(W) = 0$.*

Let $L = \sum_i L_i$ and define $T(X) = B$ as follows: $T(L_i)$ is a single point, $T$ is a homeomorphism on $X - L$. We shall see that this transformation satisfies all the conditions required of $T$. It is evident from the definition that $T$ is single-valued. Also, $T$ is an $R_0^{01}$-transformation because $T^{-1}(b)$ for every point $b$ of $B$ is a single point or some $L_i$, $L_i$ is a continuum, and $R^1(L_i) = 0$. We prove that $T$ is continuous by showing that the collection of all the sets $T^{-1}(b)$ is an upper semi-continuous collection of continua filling $X$. Since every set $T^{-1}(b)$ which is in $L$ is a component $L_i$ of $L$ and each $L_i$ is compact, it follows from a theorem of R. L. Moore [12] that the sets $T^{-1}(b)$ which are in $L$ form an upper semi-continuous collection of continua filling $L$. Furthermore, if $\{x_i\} = \{T^{-1}(b_i)\}$ is a sequence of points of $X - L$ which converges to $x$, a point of $T^{-1}(b)$, then every infinite subsequence of $\{T^{-1}(b_i)\}$ converges to $x$. Hence, all the sets $T^{-1}(b)$ form an upper semi-continuous collection of continua filling $X$. We conclude that $T$ is continuous, for Kuratowski [13] has shown that an upper semi-continuous collection of closed sets

[10] G. T. Whyburn, *American Journal of Mathematics*, vol. 50 (1928), theorem 10, p. 177.

[11] G. T. Whyburn, *American Journal of Mathematics*, vol. 56 (1934), theorem (3.1), p. 298.

[12] See footnote 7, Theorem 20, p. 340.

[13] C. Kuratowski, *Fundamenta Mathematicae*, vol. 11 (1928), theorems 2, 3, pp. 172-173.

filling a compact metric space is equivalent to a single-valued, continuous transformation defined on the space.

It is readily seen that $T(L)$ is a Moore-Kline subset of $B$.[14] The set $T(L)$ is closed and compact because $T$ is continuous. Since $T$ is also a $R_0{}^0$-transformation, connectedness is an invariant of $T^{-1}$ [15] and, hence, each component of $T(L)$ is a single point. It follows from a theorem of Zippin [16] that $T(L)$ is contained in a tree $B_1$ of $B$.

Since $T$ is continuous and monotonic, $T^{-1}(B_1)$ is a continuum. It has been shown by Vietoris [17] that the one-dimensional Betti number of a compact closed set in a metric space is an invariant of $T$. Therefore, $R^1[T^{-1}(B_1)]=0$ because $R^1(B_1)=0$. Since $T(L)\subset B_1$, $L\subset T^{-1}(B_1)$. Hence, $S\subset T^{-1}(B_1)$ and $T^{-1}(B_1)$ is the set $W$ described in the statement of the theorem.

THEOREM 6. *If $B=T(A)$ is homeomorphic to $A$ for every $T$ and every component of $\bar{P}\cdot C_i$ is contained in a continuum $L_{ij}$ of $C_i$ such that $R^1(L_{ij})=0$, where $P$ is the set of all cut points $A$ and $C_i$ is any true cyclic element of $A$, then $A\equiv\sum_i C_i$, $C_i$ is a simple closed curve, and $C_i\cdot C_j=p$, a single point, for $i\neq j$.*

It follows from Lemma 3 that the continua $L_{ij}$ of $C_i$ are contained in a continuum $L_i$ of $C_i$ such that $R^1(L_i)=0$. We shall prove first that the components $\bar{S}_j$ of $\overline{\sum_i L_i}$ also satisfy the hypothesis of Lemma 3. Let $S_1$ be any component of $\sum_i L_i$. Then $S_1=\sum_n L'_n$ where $L'_n\subset C'_n$. If $S_1=\sum_{n=1}^k L'_n$, where $k$ is finite, $S_1\equiv\bar{S}_1$. Also, $N_1=\sum_{n=1}^k C'_n$ is a Peanian continuum. If $S_1=\sum_{n=1}^\infty L'_n$, then $N_1=\sum_{n=1}^\infty C'_n$ is a connected subset of $A$ which contains only cyclic elements of $A$. Hence, by a theorem of Whyburn,[18] $\bar{N}_1$ is a continuous curve and each point of $\bar{N}_1-N_1$ is a cyclic element of $A$, namely, an end point or a cut point of $A$. Since $\bar{N}_1$ is a continuous curve in both cases, $R^1(\bar{N}_1)=\sum_n R^1(C'_n)$.[19] Hence, the basis of the 1-cycles of $\bar{N}_1$ may be considered a set of 1-cycles each of which is contained in some $C'_n$. Since $S_1\subset N_1$, it follows that $\bar{S}_1\subset\bar{N}_1$ and that any 1-cycle of $\bar{S}_1$ is a 1-cycle of $\bar{N}_1$. There-

[14] L. Zippin, *Transactions of American Mathematical Society*, vol. 34 (1932), pp. 705-721.

[15] See footnote 11, theorem (1.3), p. 295.

[16] See footnote 14, theorem 5.1, p. 709.

[17] L. Vietoris, *Mathematische Annalen*, vol. 97 (1927), theorem 8b, p. 470.

[18] See footnote 10, theorem 11, p. 177.

[19] See footnote 11, theorem 2.5, p. 138.

fore, the basis of 1-cycles of $\bar{S}_1$ can be chosen so that it is a set of 1-cycles, each of which is contained in some $C'_n$ and hence in some $L'_n$. It follows that $R^1(\bar{S}_1) = R^1(\sum_n L'_n) = 0$ because $R^1(L'_n) = 0$ for every $n$.

Furthermore, any component $\bar{S}_j$ of $\overline{\sum_i L_i}$ satisfies the same conditions as one of the sets $\bar{S}_1$ just considered. To see this we need to consider $\overline{\sum_i L_i} - \sum_i L_i$. Let $x$ be any point of this set. Then $x$ is a limit point of a subsequence $\{L_{i_j}\}$ of $\sum_i L_i$ and hence is contained in $\bar{S}_j = \overline{\sum L_{i_j}}$. Therefore, we conclude that $\overline{\sum_i L_i} = \sum_j \bar{S}_j$, where $\bar{S}_j$ is a continuum, $R^1(\bar{S}_j) = 0$, and $\bar{S}_i \cdot \bar{S}_j = 0$, $i \neq j$. Thus, the sets $\bar{S}_j$ satisfy the hypothesis of Lemma 3 and are contained in a continuum $S$ of $A$ such that $R^1(S) = 0$.

We show now that $C_i \cdot C_j = p$, a single point, for $i \neq j$. Define $T(A)$ as follows: $T(S)$ is a single point, $T$ is a homeomorphism on $A - S$. If $S$ does not cut $C_i$, $T(C_i)$ is a true cyclic element of $B$. If $S$ cuts $C_i$, let $D_{ij}$ be a component of $C_i - C_i \cdot S$. It will be demonstrated that $T(D_{ij} + C_i \cdot S) = B_{ij}$ is a true cyclic element of $B$. Since $T$ is continuous, $B_{ij}$ is a continuum. Also, $B_{ij}$ can fail to be locally connected only at the point $T(S)$ and consequently is locally connected.[20] It remains to be seen that $B_{ij}$ is cyclicly connected.

Let $x$ and $y$ represent any two points of $D_{ij}$. Since $C_i$ is cyclicly connected and $D_{ij} \subset C_i$, the points $x$ and $y$ lie on some simple closed curve $C'$ of $C_i$. If $C' \subset D_{ij}$, $T(C')$ is a simple closed curve in $B_{ij}$ because $T$ is a homeomorphism on $D_{ij}$. If $C' \not\subset D_{ij}$, there are two cases to be considered. The curve $C'$ may satisfy the condition that one of the arcs $xy$ of $C'$ is in $D_{ij}$. Then there is an arc $uxyz$ of $C'$ such that $\langle uz \rangle \subset D_{ij}$ and $u + z \subset C_i \cdot S$. Since $T(u) = T(z)$ and $T$ is a homeomorphism on $\langle uz \rangle$, $T(uxyz)$ is a simple closed curve of $B_{ij}$. Hence, in this case $T(x)$ and $T(y)$ are on a simple closed curve in $B_{ij}$.

In the second case neither arc $xy$ of $C'$ is contained in $D_{ij}$. Then there are arcs $uxv$ and $wyz$ of $C'$ such that $\langle uxv \rangle$ and $\langle wyz \rangle$ are in $D_{ij}$ and $u$, $v$, $w$, $z$ are points of $S$. There is an arc $\delta$ in $D_{ij}$ joining $x$ and $y$ because $D_{ij}$ is a component of $C_i - C_i \cdot S$. In the order $x$ to $y$ let $s$ be the last point of $uxv$ and $t$ the first point of $wyz$ on $\delta$. There are four cases depending on the location of $s$ and $t$. Suppose that $s$ is a point of $xv$ and $t$ is a point of $wy$. Then the arc $\alpha = uxs + st + tyz$, where $st \subset \delta$ and $uxs + tys \subset C'$, satisfies the conditions that $x + y \subset \alpha$, $\langle \alpha \rangle \subset D_{ij}$, and $u + v \subset S$. Hence, $T(\alpha)$ is a simple closed curve of $B_{ij}$ for the same reasons as those given in the preceding

[20] See footnote 7, theorem 8, p. 95.

case for $T(uxyz)$. The other cases corresponding to other positions of $s$ and $t$ on $uxv$ and $wyz$ can be discussed similarly. Therefore, we conclude that $B_{ij}$ is cyclicly connected.

Since $D_{ij}$ was any component of $C_i - C_i \cdot S$, it follows that $T(C_i)$ is a set of true cyclic elements of $B$ having only the point $T(S)$ in common. Furthermore, $B_i$, any true cyclic element of $B$, is contained in $T(C_j)$ for some $j$. Hence, $B_i \cdot B_j = T(S)$ for $i \neq j$ and $B$ is homeomorphic to $A$ only if $C_i \cdot C_j$ is a single point $p$, $i \neq j$.

It follows from Theorem 5 that $A \equiv \sum_i C_i$ for if some $C_i$ contains a cut point $p'$ of $A$, $p' \neq p$, the closed extension of every component of $A - C_i$ contains a true cyclic element of $A$. Hence, there are in $A$ true cyclic elements $C_j$ and $C_k$ such that $C_j \cdot C_i = p$ and $C_k \cdot C_i = p'$. It has just been shown that this is impossible since $C_i \cdot C_j = p$ for every $i \neq j$.

If some $C_i$, say $C_1$, is not a simple closed curve, let $M'$ be a manifold homeomorphic to $M$ and let $H(C_1)$ be the homeomorph of $C_1$ in $M'$. Then $H(C_1)$ is a cyclic connected Peanian continuum contained in a compact 2-manifold without boundary and hence Lemma 3 can be applied. Since the 2-cell $E$, which was used in the proof of Lemma 3, can be chosen so that $\bar{E}$ does not contain a certain point $p \neq x$, there is an acyclic cut set $A_1$ of $H(C_1)$ which does not contain $H(p)$. Define $T(A)$ thus: $T[H^{-1}(A_1)]$ is a single point, $T$ is a homeomorphism on $A - A_1$. Then $B$ contains two cut points, $T(p)$ and $T[H^{-1}(A_1)]$, and hence is not homeomorphic to $A$. We conclude that each $C_i$ is a simple closed curve.

The Peanian continuum $A \equiv \sum_i C_i$, $C_i \cdot C_j = p$ for $i \neq j$, $C_i$ a simple closed curve for every $i$, satisfies the hypothesis that $B$ is homeomorphic to $A$ for every $T$. Since $T$ is continuous and monotonic, $T^{-1}(b)$ for every point $b$ of $B$ is a continuum. Moreover, $T^{-1}(b) \cdot C_i$, for any $i$, is a continuum because the intersection of any continuum and a true cyclic element of a continuous curve is a continuum. Consequently, $T$ is monotonic on each simple closed curve $C_i$ and $T(C_i)$ is a simple closed curve by the Corollary to Theorem 1. Also, $T(C_i) \cdot T(C_j) = T(p)$, $i \neq j$, because $T$ is a $R_0{}^0$-transformation. Hence, $B$ is homeomorphic to $A$.

COROLLARY. *If $A$ contains only a finite number of true cyclic elements $C_i$, $i = 1, \cdots, n$, and $B = T(A)$ is homeomorphic to $A$ for every $T$, then $A = \sum_{i=1}^{n} C_i$, $C_i \cdot C_j = p$, $i \neq j$, $C_i$ is a simple closed curve.*

UNIVERSITY OF MICHIGAN.

# ON THE USE OF INDETERMINATES IN THE THEORY OF THE ORTHOGONAL AND SYMPLECTIC GROUPS.*

By HERMANN WEYL.

**1.** These lines contain a supplement, concerning the use of indeterminates, to my book, *The Classical Groups*, Princeton (1939), which I cite in the following as *CG*. The general infinitesimal operation of the orthogonal group,

$$dx = Sx, \tag{1}$$

is described by a skew-symmetric matrix $S = \| s_{ik} \|$ of which the $n(n-1)/2$ elements $s_{ik}$ $(i < k)$ are indeterminate parameters. It is to be expected that the first main theorem for vector invariants of the orthogonal group will hold if invariance is demanded only with respect to the infinitesimal operations (1). In this form the theorem cannot be affected by the field in which we operate, as long as it is of characteristic zero. In my book I approached the problem of vector invariants for all groups discussed in a systematic way by combining the formal apparatus of Capelli's identities with such non-formal reasoning as is used to prove the geometric theorem of congruence. The topological fact that the proper orthogonal group is a connected manifold shows that for the field $\mathbf{K}$ of all real numbers the full group may be replaced by the set of its infinitesimal elements. Actually I employed an algebraic equivalent for this topological argument which, by a somewhat devious procedure, carries the results over to "formal" and "infinitesimal" invariants. I now think it would have been wiser to settle the question of infinitesimal invariants for the orthogonal and the symplectic groups directly by means of another formal identity of Capelli's type which I propose to develop in § 2.

Besides the question of invariants there is connected with any group $\Gamma$ of linear transformations $A = \| a_{ik} \|$ in a vector space $\mathbf{P}$ an ideal $\mathfrak{J}(\Gamma)$. It consists of all polynomials $\Phi(A)$ of the $n^2$ variables $a_{ik}$ which vanish for each element $A$ of $\Gamma$. Whereas invariance under the whole group is secured when it holds for any set of generators of the group, e. g., for the infinitesimal elements, it is essential to postulate the vanishing of $\Phi$ for *all* elements $A$ of the group; for it is not true that the vanishing of $\Phi(A_1)$ and $\Phi(A_2)$ implies the vanishing of $\Phi(A_1 A_2)$. In proving that $(E - S)/(E + S)$ is a generic zero of the orthogonal ideal, I should therefore have observed that the

* Received May 16, 1941.

expression $(E-S)/(E+S)$ with its $n(n-1)/2$ indeterminates $s_{ik}$ $(i<k)$ constitutes a group, in the sense that the equation

$$\frac{E-S}{E+S}=\frac{E-S'}{E+S'}\cdot\frac{E-S''}{E+S''} \tag{2}$$

defines a skew-symmetric matrix $S$ in terms of two indeterminate skew-symmetric matrices $S'$ and $S''$; the parameters of $S$ are rational functions of those of $S'$ and $S''$. Not only did I fail to emphasize this point, but the wrong formulation of Theorem (5. 3. B) betrays that I overlooked it altogether. The elements

$$\frac{E-S}{E+S}, \qquad J_n\cdot\frac{E-S}{E+S} \tag{3}$$

constitute a group in the above sense. Hence the hypothesis imposed upon $\Phi(A)$ in that theorem should have required its vanishing for (3) (and not merely for $(E-S)/(E+S)$ and $J_n$). The corollary immediately following the theorem and Lemma (7. 1. B) ought to be corrected accordingly. (§ 3)

An important link of the investigation is the fact that the set $\Pi_f[(E-S)/(E+S)]$, being a set of orthogonal transformations in the tensor space $P_f$, is fully reducible. We shall deal with this point without making numerical substitutions for the indeterminates $s_{ik}$ and, in the case of the symplectic group, without recourse to the "unitarian trick." (§ 4) We carry the details through for the symplectic group. The modifications required for the orthogonal group are obvious.

**2.** The coördinates of a vector $x$ in $n=2\nu$ dimensions will be denoted by

$$x_1,\cdots,x_\nu;\ x_{1'}=x'_1,\cdots,x_{\nu'}=x'_\nu.$$

Indices $i,k,\cdots$ run over the whole range $1,\cdots,\nu,\ 1',\cdots,\nu'$ whereas $\alpha,\beta,\cdots$ run merely over the half range $1,\cdots,\nu$. The skew-product $[xy]$ of two vectors $x,y$ is defined by

$$[xy]=\sum_a(x_ay'_a-x'_ay_a)=\sum_{i,k}\epsilon(ik)x_iy_k.$$

Its matrix $\|\epsilon(ik)\|$ is called $I$. An infinitesimal operation $dx=Sx$ of the symplectic group has a matrix

$$S=\begin{Vmatrix} s(\alpha\beta), & s(\alpha\beta') \\ s(\alpha'\beta), & s(\alpha'\beta') \end{Vmatrix}$$

for which

$$s(\alpha'\beta')=-s(\beta\alpha);\quad s(\alpha\beta')=s(\beta\alpha'),\quad s(\alpha'\beta)=s(\beta'\alpha). \tag{4}$$

As its $n(n+1)/2$ parameters $s_j$ we use

$$\text{all } s(\alpha\beta);\quad s(\alpha\beta') \text{ for } \alpha\leqq\beta;\quad s(\alpha'\beta) \text{ for } \alpha\leqq\beta.$$

We refer to these conditions (4) by speaking of an $S$-matrix. Thus we write the indeterminate $S$-matrix as $L(s) = \Sigma L_j s_j$.

Let $f(x, y, \cdots)$ be a polynomial homogeneous of a certain degree with respect to the components of each argument vector. Let $e$ be the number of argument vectors and $r$ the degree of $f$ with respect to $x$. In order to express invariance of $f$ under infinitesimal symplectic transformations, we introduce the matrix $X = x\xi$ with the elements $X(ik) = x_i\xi_k$, composed of a covariant vector $x$ (single column) and a contravariant $\xi$ (single row), and moreover the matrix $X'$ with the elements

$$X'(\alpha\beta) = x_\alpha\xi_\beta - x'_\beta\xi'_\alpha, \qquad X'(\alpha\beta') = x_\alpha\xi'_\beta + x_\beta\xi'_\alpha,$$
$$X'(\alpha'\beta) = x'_\alpha\xi_\beta + x'_\beta\xi_\alpha, \qquad X'(\alpha'\beta') = x'_\alpha\xi'_\beta - x_\beta\xi_\alpha.$$

Incidentally

$$X' = X + IX^*I.$$

Setting $\xi_i = \partial/\partial x_i$ turns $X(ik)$ and $X'(ik)$ into differential operators $d_x(ik)$, $d'_x(ik)$. The infinitesimal invariance of $f$ is expressed by

$$\sum_y d'_y(ik)f = 0,$$

the sum extending over the $e$ argument vectors $y$ of $f$.

Let us start with the identity

$$\begin{vmatrix} [xy] & (x\eta) \\ (\xi y) & [\xi\eta] \end{vmatrix} = -\operatorname{tr}(XY'). \tag{5}$$

(If one prefers, one may write the trace on the right side as $\frac{1}{2}\operatorname{tr}(X'Y')$, but calculations are a little bit easier with the original form.) Put $\xi = \partial/\partial x$, $\eta = \partial/\partial y$; the polar $D_{yx}f$ and the operator

$$\Theta_{[xy]}f = \sum_a \left(\frac{\partial^2}{\partial x_a \partial y'_a} - \frac{\partial^2}{\partial x'_a \partial y_a}\right) f$$

make their appearance. Both are infinitesimal symplectic invariants if $f$ is. $\Theta_{[xy]}$ plays the same rôle for the symplectic group as the Laplacian operator

$$\Theta_{xx}f = \sum_i \frac{\partial^2 f}{\partial x_i^2}$$

does for the orthogonal group. Handling the operators $D$ and $d$ as if multiplication by $x_i$ and differentiation $\partial/\partial x_i$ were commutable, one would obtain from (5) the equation

$$[xy] \cdot \Theta_{[xy]}f - D_{xy}(D_{yx}f) = -\operatorname{tr}(d_x, d'_y f).$$

What one really gets are the equations

$$[xy] \cdot \Theta_{[xy]} f - D_{xy}(D_{yx} f) + rf = -\operatorname{tr}(d_x, d'_y f)$$

if $y \neq x$, and

$$[xx] \cdot \Theta_{[xx]} f - r(r+n) f = -\operatorname{tr}(d_x, d'_x f)$$

for $y = x$. Under the assumption that $f$ is an infinitesimal symplectic invariant (briefly, an invariant), we thus find

$$\sum_y [xy] \Theta_{[xy]} f - \sum_y{}' D_{xy}(D_{yx} f) = r(r+n+1-e) f \tag{6}$$

where $\sum_y, \sum_y{}'$ designate summations over all $e$ arguments $y$ including or excluding the fixed argument $x$.

We maintain:

THEOREM I. *Every invariant $f(x, y, \cdots)$ is expressible as a polynomial in terms of the skew-products $[xy], \cdots$ of its arguments $x, y, \cdots$.*

Suppose first $e \leqq n+1$. The polar $D_{yx} f$ for $y \neq x$ and $\Theta_{[xy]} f$ (whether $y = x$ or $y \neq x$) are invariants of lower degree in $x$ than $f$. Hence the equation (6) proves the theorem by induction with respect to $r$ for any degree $r$ under the hypothesis that it is true for $r = 0$, i. e., for invariants depending only on the $e - 1$ arguments $y, \cdots$. Thus induction with respect to $e$ yields the desired result for $e \leqq n+1$. But Capelli's general identity allows us to increase the number of arguments indefinitely once the stage of $n$ arguments is reached. (There is an overlapping inasmuch as the increase of $e$ from $n$ to $n+1$ may be effected either by that identity or by (6). Up to $n+1$ arguments there are no algebraic relations between the skew-products of the arguments.)

For the orthogonal group (6) is replaced by the formula

$$\sum_y (xy) \Theta_{xy} f - \sum_y{}' D_{xy}(D_{yx} f) = r(r+n-1-e) f.$$

It works for $e \leqq n-1$ while Capelli's general identity goes on from $e = n$. The gap $e = n-1 \rightarrow n$ is bridged by Capelli's special identity which brings in the "bracket factor."

**3.** The relationship

$$A = (E-S)(E+S)^{-1} = (E+S)^{-1}(E-S) \tag{7}$$

between finite and infinitesimal symplectic transformations, $A$ and $S$, shows that (2) defines an $S$-matrix in terms of two indeterminate $S$-matrices $S', S''$ with the parameters $s_j', s_j''$. The parameters $s_j$ of $S$ are rational functions $\psi_j(s', s'')$. Writing (2) in the form

$$2(E+S)^{-1}-E=(E+S')^{-1}(E-S')(E-S'')(E+S'')^{-1}$$

one finds at once

$$(E+S)^{-1}=(E+S')^{-1}(E+S'S'')(E+S'')^{-1}$$

or

$$E+S=(E+S'')(E+S'S'')^{-1}(E+S').$$

Hence the functions $\psi_j(s',s'')$ have the determinant

$$\Delta=|E+S'S''|$$

as their common denominator, and

$$(8) \qquad |E+S|=|E+S'|\cdot|E+S''|\cdot\Delta^{-1}.$$

We derive the product matrix $\Pi_f A$ in the tensor space $P_f$ from

$$(9) \qquad A=(E-S)(E+S)^{-1}, \qquad S=L(s).$$

Evidently it is of the following form

$$(10) \qquad \Pi_f A=R(s)/|E+S|^f, \qquad R(s)=\sum_p C_p\sigma_p$$

where the numerator $R(s)$ is a polynomial of formal degree $nf$ of the parameters $s_j$ with matric coefficients $C_p$. Thus $\sigma_p$ runs over the monomials of degree $\leqq nf$ of the parameters $s_j$. The same holds for $\Pi^{(f)}(A)$:

$$|E+S|^f\cdot\Pi^{(f)}(A)=\sum_p C_p^{(f)}\sigma_p.$$

We maintain

THEOREM II. *The linear closure of* (10), *i. e., the set* $\mathfrak{C}_f$ *of all linear combinations* $\Sigma\lambda_pC_p$ *with numerical coefficients* $\lambda_p$, *is an algebra containing the unit matrix.*

The unit matrix is the coefficient $C_0$ of the monomial $\sigma_0=1$. Equation (2) entails

$$\Pi_f\left(\frac{E-S}{E+S}\right)=\Pi_f\left(\frac{E-S'}{E+S'}\right)\cdot\Pi_f\left(\frac{E-S''}{E+S''}\right)$$

or, because of (8),

$$\Delta^f\cdot\sum_r C_r\sigma_r=\sum_{p,q}C_pC_q\sigma_p'\sigma_q''.$$

After the substitution of $\psi_j(s',s'')$ for $s_j$ each $\sigma_r$ turns into a polynomial $\phi_r(s',s'')$ divided by $\Delta^{nf}$. Hence

$$(11) \qquad \Delta^{(n-1)f}\cdot\sum_{p,q}C_pC_q\sigma_p'\sigma_q''=\sum_r C_r\phi_r(s',s'').$$

We now apply the following trivial algebraic lemma:

Let $\phi = 1 + \cdots$ be a given polynomial of some variables $x_1, \cdots, x_l$, with the constant term 1. Then the coefficients of an arbitrary polynomial $F$ of degree $m$ may be linearly expressed by the coefficients of $\phi F = G$.

Arranging the terms of $F$ in ascending lexicographic order one obtains recursive linear equations for the unknown coefficients $a$ of $F$, with the coefficients $b$ of $G$ as the known right members (only terms of $G$ of a degree not exceeding $m$ enter). Denote by $F_\mu$ the terms of degree $\mu$ in any polynomial $F = F_0 + F_1 + \cdots$. Putting $\phi = 1 - \omega$ and using the power series $(1 - \omega)^{-1}$, one finds more explicitly

$$F = \Sigma\, G_\mu\, \omega_{\mu_1}\, \omega_{\mu_2} \cdots$$
$$(\mu \geqq 0;\ \mu_1 \geqq 1, \mu_2 \geqq 1, \cdots;\ \mu + \mu_1 + \mu_2 + \cdots \leqq m).$$

If $\phi$ has integral coefficients, then the coefficients of the linear combinations expressing the $a$ in terms of the $b$ are likewise integers.

This lemma is of immediate application to the equation (11) in which $\Delta$ begins with the constant term 1. The products $\sigma_p'\sigma_q''$ are monomials no two of which are equal, and thus we arrive at equations

$$C_p C_q = \sum_r \gamma^r{}_{pq} C_r$$

with rational integers $\gamma^r{}_{pq}$. They prove our statement. The same equations hold for the matrices $C_p^{(f)}$, and thus the linear combinations $\Sigma \lambda_p C_p^{(f)}$ form an algebra $\mathfrak{C}^{(f)}$. The stage for the whole drama is the field $\kappa$ of rational numbers (or, as the arithmetician might care to observe, the even narrower ring of rational integers).

**4.** We go on operating in $\kappa$. The next step is

THEOREM III. *The algebra $\mathfrak{C}_f$ in $\kappa$ is fully reducible.*

Suppose that a subspace $\{a_\lambda\}$ of $\mathrm{P}_f$ is spanned by a number of linearly independent vectors

$$a_\lambda = (a_{1\lambda}, \cdots, a_{N\lambda}), \qquad (N = n^f).$$

The orthogonal subspace $\{a_\mu\}$ is spanned by a complete set of linearly independent solutions $x = a_\mu$ of the simultaneous equations

$$\sum_{i=1}^{N} a_{i\lambda} x_i = 0.$$

(The ranges of the two indices $\lambda$ and $\mu$ are disjunct.) In $\kappa$ the two subspaces have no vector in common except 0 and hence span the whole space; or the matrix $U$ whose columns are the $a_\lambda$ and $a_\mu$ is a non-singular square matrix. By construction the symmetric matrix $U^*U$ decomposes into a $(\lambda, \lambda)$- and a

$(\mu, \mu)$-part whereas the rectangles $(\lambda, \mu)$ and $(\mu, \lambda)$ are empty. We now assume the first subspace $\{a_\lambda\}$ to be invariant under $\mathfrak{C}_f$. This circumstance is expressed by an equation

$$R(s) \cdot U = U \cdot Q(s)$$

where $Q(s)$ is reduced in the sense that its $(\lambda, \mu)$-part is empty. We propose to prove that its $(\mu, \lambda)$-part is also empty.

To that end, multiply by $U^*$ to the left:

$$U^*R(s)U = (U^*U)Q(s).$$

We then see that the $(\lambda, \mu)$-part of $U^*R(s)U$ is empty,

$$[U^*R(s)U]_{(\lambda,\mu)} = 0, \tag{12}$$

and that it is sufficient to establish the same fact for the $(\mu, \lambda)$-part of the same matrix.

If $S$ is an $S$-matrix, so is its transpose $S^*$. The following simple linear involutorial substitution of the parameters $s_j$,

$$s(\alpha\beta) \to s(\beta\alpha), \quad s(\alpha\beta') \to s(\alpha'\beta), \quad s(\alpha'\beta) \to s(\alpha\beta'), \tag{13}$$

changes the indeterminate $S = L(s)$ into $S^*$, hence $A$, (9), into $A^*$, $\Pi_f A$ into $(\Pi_f A)^*$, and since $|E + S| = |E + S^*|$, also $R(s)$ into $R^*(s)$ and $R^*(s)$ back into $R(s)$. Taking the transpose of the equation (12),

$$[U^*R^*(s)U]_{(\mu,\lambda)} = 0,$$

and then carrying out the substitution (13), one arrives at the desired result

$$[U^*R(s)U]_{(\mu,\lambda)} = 0.$$

(13) induces a simple involutorial permutation among the monomials $\sigma_p$, $\sigma_p \to \sigma_{p^*}$. Our argument shows that

$$\Sigma C^*{}_p\, \sigma_p = \Sigma C_p\, \sigma_{p^*} = \Sigma C_{p^*}\, \sigma_p; \quad C^*{}_p = C_{p^*}.$$

Thus the algebra $\mathfrak{C}_f$ is a set $\mathfrak{C}$ of matrices $C$ which coincides with the set $\mathfrak{C}^*$ of its transposed elements $C^*$. The method employed yields this general proposition: *For any set $\mathfrak{C}$ of matrices in a real field $k$ with the property $\mathfrak{C}^* = \mathfrak{C}$ reduction results in full reduction if projection modulo the invariant subspace is carried out as projection upon the orthogonal subspace.* This is in line with older investigations by E. Fischer who studied groups $\mathfrak{C}$ of linear transformations in a real field $k$ enjoying the property $\mathfrak{C}^* = \mathfrak{C}$ and showed that their invariants depending on a variable form have a finite integrity basis.[1]

Once in possession of our three theorems we put the same machinery into

[1] *Journal für reine u. ang. Mathematik*, vol. 140 (1911), pp. 48-81.

play as employed on pp. 141-142 of *CG*. We still operate in the field $\kappa$. An $f(x, y, \cdots)$ which is invariant under the generic (9) is necessarily an infinitesimal invariant. Apply the hypothesis to $tS$ instead of $S$ and take merely the first power of the parameter $t$ into account. Consequently Theorem I shows that the algebra described as $\mathfrak{A}^{(f)}$ on p. 174 is the commutator algebra of the commutator algebra of $\Pi^{(f)}[(E-S)/(E+S)]$ or of $\mathfrak{C}^{(f)}$. Hence as $\mathfrak{C}^{(f)}$ is an algebra containing the unit matrix (Theorem II) and fully reducible (Theorem III), the algebra $\mathfrak{A}^{(f)}$ is not wider than $\mathfrak{C}^{(f)}$ (R. Brauer's principle). In view of the definition of $\mathfrak{A}^{(f)}$ this fact remains true in any field of characteristic zero. Any polynomial $\Phi(A)$ of degree $f$ depending on the matrix $A = \| a_{ik} \|$ with variable components $a_{ik}$ proceeds from a linear form $\gamma(A^{(f)})$ of an arbitrary bisymmetric $A^{(f)}$ by the substitution $A^{(f)} = \Pi^{(f)}(A)$. If $\Phi(A)$ is annulled by the substitution (9) we must have $\gamma(C_p{}^{(f)}) = 0$, and therefore $\gamma(A^{(f)})$ vanishes for all matrices $A^{(f)}$ of $\mathfrak{A}^{(f)}$. Thus results Theorem (6. 3. B) on p. 174 of *CG*. In restating it I propose the following terminology for ideals of polynomials $\phi(x_1, \cdots, x_l)$. Elements $\phi_1, \cdots, \phi_m$ of a given ideal $\mathfrak{J}$, whose degrees are $f_1, \cdots, f_m$ respectively, are said to constitute a *form basis* of $\mathfrak{J}$ if any element $\phi$ of $\mathfrak{J}$ of degree $f$ may be obtained in the form

$$\phi = h_1\phi_1 + \cdots + h_m\phi_m$$

*where $h_i$ is a polynomial of degree $f - f_i$* (which implies, of course, the vanishing of $h_i$ whenever $f - f_i < 0$). By introduction of homogeneous coördinates $\mathfrak{J}$ gives rise to an ideal $\bar{\mathfrak{J}}$ of homogeneous forms $\phi(x_0, x_1, \cdots, x_l)$ such that $\phi(x_0, x_1, \cdots, x_l)$ is in $\bar{\mathfrak{J}}$ if and only if $\phi(1, x_1, \cdots, x_l)$ is in $\mathfrak{J}$. The statement "$\phi_1, \cdots, \phi_m$ constitute a *form basis* of $\mathfrak{J}$" means that the corresponding homogeneous forms $\phi_1, \cdots, \phi_m$ of degrees $f_1, \cdots, f_m$ constitute a *basis* of $\bar{\mathfrak{J}}$.

We arrive at the following proposition concerning the *symplectic ideal* (i. e., the ideal $\mathfrak{J}(\Gamma)$ of the symplectic group $\Gamma$) which holds in any field $k$ of characteristic 0:

THEOREM IV. 1) *Let $A = \| a_{ik} \|$ be the matrix with $n^2$ variable components $a_{ik}$. The components of the two matrices $A^*IA - I$ and $AIA^* - I$ constitute a form basis, the components of either of them a basis, of the symplectic ideal.*

2) *The symplectic ideal is a prime ideal, and the expression* (9) *a generic zero; i. e., a polynomial $\phi(A)$ vanishes for all symplectic transformations $A$ if it is annulled by the substitution* (9).

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, N. J.

# AN ALGEBRAIC PROOF OF A PROPERTY OF LIE GROUPS.*

By Claude Chevalley.

Let $\mathfrak{G}$ be a Lie group with complex parameters, and let $\mathfrak{L}$ be its Lie algebra. To every element $X \epsilon \mathfrak{L}$ there corresponds an infinitesimal transformation of the adjoint group, which we shall denote by $\mathbf{X}$; $\mathbf{X}$ is a linear endomorphism of $\mathfrak{L}$, defined by the formula

$$\mathbf{X}Y = [Y, X] \text{ for every } Y \epsilon \mathfrak{L}.$$

If we select a base in $\mathfrak{L}$, every $\mathbf{X}$ may be represented by a matrix. We denote by $l$ the smallest integer such that $\mathfrak{L}$ contains elements $X$ for which 0 is an $l$-uple characteristic root of $\mathbf{X}$. Such elements $X$ are said to be *regular.*

E. Cartan has shown [1] that every regular element $X$ belongs to a uniquely determined nilpotent sub-algebra [2] $\mathfrak{N}$ of dimension $l$ of $\mathfrak{L}$. The algebra $\mathfrak{N}$ is composed of those elements $Y \epsilon \mathfrak{L}$ for which $\mathbf{X}^l Y = 0$. We propose to call such a nilpotent sub-algebra a *Cartan algebra* of the Lie algebra $\mathfrak{L}$.

Any automorphism of $\mathfrak{L}$ permutes among themselves the Cartan algebras of $\mathfrak{L}$. This applies, in particular, to the operations of the adjoint group. We want to prove the following result:

Theorem. *If $\mathfrak{N}_0$ and $\mathfrak{N}_1$ are any two Cartan algebras of the Lie algebra $\mathfrak{L}$, there always exists an operation of the adjoint group which carries $\mathfrak{N}_0$ onto $\mathfrak{N}_1$.*

This theorem is well known in the case where $\mathfrak{L}$ is semi-simple.[3] It turns out that the proof, when translated in algebraic terminology, may be extended so as to include the general case.

The possibility of doing this is somehow surprising at first sight, since the adjoint group itself is not necessarily algebraic. However, the success of

* Received May 22, 1941.

[1] Cf. Cartan, Thèse, Paris, 1894, or Zassenhaus, "Über Lie'sche Ringe mit Primzahlcharakteristik," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 13 (1939), pp. 1-100.

[2] The nilpoint algebras are also called "groups of rank 0."

[3] Cf. H. Weyl, "Theorie der Darstellung kontinuierlicher halb-einfacher gruppen durch lineare Transformationen, part III," *Mathematische Zeitschrift*, vol. 24 (1925-26), pp. 377-395. For a proof of the theorem in the case of compact Lie groups, cf. André Weil, "Démonstration topologique d'un théorème fondamental de Cartan," *Comptes Rendus de l'Académie des Sciences*, vol. 200 (1935), p. 518. Although his proof is based on topological considerations, it is in some ways related to the proof which is given here.

our method depends on the fact that the adjoint group contains an algebraic sub-group which has already sufficiently many elements to transform any given Cartan algebra into any other.

Before entering into the details, we remark that the Cartan algebras play in the theory of Lie algebras the same rôle that the Sylow sub-groups play in the theory of finite groups. Our theorem corresponds to the fact that the Sylow sub-groups of a finite group (corresponding to a prime $p$) are all conjugate sub-groups.

**I. A lemma of algebraic geometry.** Let $\Sigma$ be a field of algebraic functions over an algebraically closed field $\Omega$. Let us select any finite number $r$ of functions $\xi_1, \xi_2, \cdots, \xi_r$ in $\Sigma$; we build the algebraic manifold $\mathfrak{V}$ whose "generic point"[4] is $(\xi_1, \xi_2, \cdots, \xi_r)$. Such a manifold will always be called a "model" of $\Sigma$, even in the case where $\Omega(\xi_1, \xi_2, \cdots, \xi_r)$ is a proper sub-field of $\Sigma$. If we introduce $s$ new functions $\eta_1, \eta_2, \cdots, \eta_s$ of $\Sigma$, we can construct a new model $\mathfrak{V}_1$ of $\Sigma$ whose generic point is $(\xi_1, \xi_2, \cdots, \xi_r; \eta_1, \eta_2, \cdots, \eta_s)$. If $(\bar{\xi}; \bar{\eta})$ is any point of $\mathfrak{V}_1$, the point $(\bar{\xi})$ is clearly a point of $\mathfrak{V}$, which we call the projection on $\mathfrak{V}$ of the point $(\bar{\xi}, \bar{\eta})$.

LEMMA 1. *Let there be given a finite system $\mathfrak{M}_1$ of proper algebraic sub-manifolds of $\mathfrak{V}_1$. There exists a finite system $\mathfrak{M}$ of proper algebraic sub-manifolds of $\mathfrak{V}$ such that every point of $\mathfrak{V}$ which does not belong to $\mathfrak{M}$ is the projection of a point of $\mathfrak{V}_1$ not belonging to $\mathfrak{M}_1$.*

We shall first consider the case where $s = 1$, and we shall set $\eta_1 = \eta$. There exists, by hypothesis, a polynomial $P(x_1, x_2, \cdots, x_r; y)$ for which $P(\xi_1, \xi_2, \cdots, \xi_r; \eta) \neq 0$, but such that $P$ vanishes at every point of $\mathfrak{M}_1$. Let us consider separately two cases.

1. The function $\eta$ is algebraic over $\Omega(\xi_1, \xi_2, \cdots, \xi_r)$. Then there exists an irreducible polynomial $\phi(x_1, x_2, \cdots, x_r; y)$ which actually contains $y$ and is such that $\phi(\xi_1, \xi_2, \cdots, \xi_r; \eta) = 0$. Moreover, if $\psi(x_1, x_2, \cdots, x_r; y)$ is any polynomial which vanishes on $\mathfrak{V}_1$, $\psi(\xi_1, \xi_2, \cdots, \xi_r; y)$, considered as a polynomial with coefficients in $\Omega(\xi_1, \cdots, \xi_r)$, is divisible by $\phi(\xi_1, \xi_2, \cdots, \xi_r; y)$. It follows that there exists a polynomial $\Theta(x_1, \cdots, x_r)$ for which

$$\Theta(\xi_1, \xi_2, \cdots, \xi_r) \neq 0$$

and such that

[4] For the notions of an algebraic manifold and of a generic point, cf. v. d. Waerden, *Vorlesungen über algebraische Geometrie* (Springer).

$$(1) \qquad \Theta(\xi_1, \xi_2, \cdots, \xi_r)\psi(\xi_1, \xi_2, \cdots, \xi_r; y) \\ = \phi(\xi_1, \xi_2, \cdots, \xi_r; y) A(\xi_1, \cdots, \xi_r; y)$$

where $A(x_1, x_2, \cdots, x_r; y)$ is a polynomial.

The variety $\mathcal{V}_1$ is determined by a prime ideal of polynomials in $x, y$; we select a base $\psi_1, \psi_2, \cdots, \psi_h$ of this ideal, and, for each one of the polynomials $\psi_i$, we write an equation of the type (1). Let $\Theta_1(x), \Theta_2(x), \cdots, \Theta_h(x)$ be the polynomials which occur in the left-hand sides of these equations.

The polynomial $P(\xi_1, \cdots, \xi_r; y)$ cannot be divisible by $\phi(\xi_1, \cdots, \xi_r, y)$, because we would otherwise have $P(\xi_1, \cdots, \xi_r; \eta) = 0$. Since $\phi$ is irreducible, there exist polynomials

$$A(x_1, \cdots, x_r, y), \quad B(x_1, x_2, \cdots, x_r, y), \quad C(x_1, \cdots, x_r)$$

such that

$$AP + B\phi = C, \quad C(\xi_1, \cdots, \xi_r) \neq 0.$$

We set $\Theta_{h+1} = C$, and we denote by $\Theta_{h+2}(\xi)$ the coefficient of the highest power of $y$ in $\phi(\xi; y)$. Let $\Theta(x)$ be the polynomial $\prod_{i=1}^{h+2} \Theta_i(x)$. Then we have $\Theta(\xi_1, \cdots, \xi_r) \neq 0$ so that the equation $\Theta = 0$ determines on $\mathcal{V}$ a finite system $\mathcal{M}$ of proper sub-manifolds. Let $(\bar{\xi}_1, \cdots, \bar{\xi}_r)$ be a point of $\mathcal{V} - \mathcal{M}$. Since $\Theta_{h+2}(\bar{\xi}) \neq 0$, the equation $\phi(\bar{\xi}, y) = 0$ has at least one solution $\bar{\eta}$. Since $\Theta_i(\bar{\xi}) \neq 0$ $(1 \leqq i \leqq h)$, the equations (1) show that $\psi_i(\bar{\xi}_1, \cdots, \bar{\xi}_r, \bar{\eta}) = 0$ $(1 \leqq i \leqq h)$; hence the point $(\bar{\xi}_1, \cdots, \bar{\xi}_r, \bar{\eta})$ belongs to $\mathcal{V}_1$. Since

$$C(\bar{\xi}) = \Theta_{h+1}(\bar{\xi}) \neq 0, \text{ and } \phi(\bar{\xi}, \bar{\eta}) = 0,$$

we have $P(\bar{\xi}; \bar{\eta}) \neq 0$, and the point $(\bar{\xi}; \bar{\eta})$ does not belong to $\mathcal{M}_1$: the lemma is proved in this case.

2. Suppose now that $\eta$ is transcendant over $\Omega(\xi_1, \cdots, \xi_r)$. Let $\Theta(\xi_1, \cdots, \xi_r)$ be the coefficient of the highest power of $y$ in $P(\xi; y)$. We have $\Theta(\xi_1, \cdots, \xi_r) \neq 0$; hence the equation $\Theta = 0$ determines a finite system $\mathcal{M}$ of proper sub-manifolds of $\mathcal{V}$. If $(\bar{\xi}_1, \cdots, \bar{\xi}_r)$ is a point of $\mathcal{V} - \mathcal{M}$, there exists a number $\bar{\eta}$ such that $P(\bar{\xi}, \bar{\eta}) \neq 0$, and the point $(\bar{\xi}, \bar{\eta})$ belongs to $\mathcal{V}_1 - \mathcal{M}_1$. Therefore the lemma is entirely proved for $s = 1$.

If $s > 1$, we denote by $\mathcal{V}_{s-i+1}$ the model whose generic point is

$$(\xi_1, \cdots, \xi_r, \eta_1, \cdots, \eta_i);$$

we have $\mathcal{V}_{s+1} = \mathcal{V}$, and, for $i = s$, we find the model already denoted by $\mathcal{V}_1$. Moreover $\mathcal{V}_{i-1}$ is obtained from $\mathcal{V}_i$ by adjoining a single new coördinate. Hence we can determine step by step in each $\mathcal{V}_i$ a finite system $\mathcal{M}_i$

of proper sub-manifolds such that $\mathfrak{M}_1$ is the given system and that every point of $\mathfrak{V}_i - \mathfrak{M}_i$ is the projection of a point of $\mathfrak{V}_{i-1} - \mathfrak{M}_{i-1}$. It is sufficient to take $\mathfrak{M} = \mathfrak{M}_{s+1}$, which completes the proof of the lemma.

*Remark.* This lemma is very closely related to the theory of relation-true specializations of v. d. Waerden.

**II. The manifold of the Cartan algebras.** We consider a Lie algebra $\mathfrak{L}$ over an algebraically closed field $\Omega$, and we select a base $\{L_1, L_2, \cdots, L_n\}$ of $\mathfrak{L}$ over $\Omega$. Let $u_1, u_2, \cdots, u_n$ be $n$ indeterminates, algebraically independent over $\Omega$. We construct the field $\Omega(u)$ obtained by adjunction of the $u_i$'s to $\Omega$. The linear combinations of the elements of $\mathfrak{L}$ with coefficients in $\Omega(u)$ form a Lie algebra $\mathfrak{L}_u$. We denote by $L_u$ the element $\Sigma u_i L_i \in \mathfrak{L}_u$, which we may consider as a generic element of $\mathfrak{L}$. As an element of $\mathfrak{L}_u$, it is obviously regular. We denote by $\mathbf{L}_u$ the matrix which represents $L_u$ in the adjoint representation of $\mathfrak{L}_u$; hence we have

$$[L_i, L_u] = \mathbf{L}_u L_i = \Sigma_j \lambda_{ji}(u) L_j, \qquad (1 \leqq i \leqq n),$$

where the $\lambda_{ji}(u)$ are linear forms whose coefficients are the constants of structure of $\mathfrak{L}$. The matrix $\mathbf{L}_u$ has exactly $l$ characteristic roots equal to 0. Therefore, we can select $l$ linearly independent elements $M_{u,1}, \cdots, M_{u,l}$ of $\mathfrak{L}_u$ such that $\mathbf{L}_u{}^l M_{u,i} = 0$ $(1 \leqq i \leqq l)$; these elements form a base of a Cartan algebra $\mathfrak{N}_u$ of $\mathfrak{L}_u$, which is the only Cartan algebra containing $L_u$. We may call $\mathfrak{N}_u$ a generic Cartan algebra of $\mathfrak{L}$. The coefficients of $\mathbf{L}_u{}^l$ being homogeneous polynomials of degree $l$ in the parameters $u$, we may assume that the coördinates of the elements $M_{u,i}$ are homogeneous functions of the $u$'s.

We shall now introduce the Plückerian coördinates of an $l$-dimensional vector sub-space $\mathfrak{M}$ of $\mathfrak{L}$. Let $M_i = \Sigma_j m_{ji} L_j$ $(1 \leqq i \leqq l)$ constitute a base for $\mathfrak{M}$. We construct over $\mathfrak{L}$ a Grassmann algebra, in which we denote the operation of multiplication by the sign $\square$. We can write

$$M_1 \square M_2 \square \cdots \square M_l = \Sigma_\mu \xi_\mu V_\mu$$

where $\mu$ stands for any arrangement $\{i_1, i_2, \cdots, i_l\}$ of $l$ of the indices $1, 2, \cdots, n$, and $V_\mu = L_{i_1} \square L_{i_2} \square \cdots \square L_{i_l}$. If we furthermore require that $\xi_\mu$ shall be a skew-symmetric function of the indices which enter in $\mu$, the coefficients $\xi_\mu$ are entirely determined when $M_1, M_2, \cdots, M_l$ are given. Moreover, it is well known that they are all multiplied by a same factor $\neq 0$ if we replace the base $\{M_1, M_2, \cdots, M_l\}$ by any other base of $\mathfrak{M}$. Hence the numbers $\xi_\mu$ may be considered as giving a system of homogeneous coördinates of $\mathfrak{M}$: they are called the Plückerian coördinates of $\mathfrak{M}$.

Let us now form the Plückerian coördinates $\Xi_\mu(u)$ of the generic Cartan algebra $\mathfrak{N}_u$. We may consider them as the coördinates of a generic point of an irreducible projective algebraic manifold. This manifold shall be denoted by $[\mathfrak{N}]$, and we shall call it the *manifold of the Cartan algebras.*

LEMMA 2. *If $\mathfrak{N}_0$ is any Cartan algebra of $\mathfrak{L}$, the Plückerian coördinates of $\mathfrak{N}_0$ are the coördinates of a point of $[\mathfrak{N}]$.*

In fact, let us select in $\mathfrak{N}_0$ a regular element $L_0 = \Sigma u_i{}^0 L_i$. The matrix $(\boldsymbol{L}_0)^l$ has exactly $l$ characteristic roots equal to zero. Therefore, it has a minor $\Delta_0$ of degree $n - l$ which is $\neq 0$. Let $\Delta(u)$ be the minor of $\boldsymbol{L}_u{}^l$ formed of the same lines and columns as $\Delta_0$. We have $\Delta(u_0) \neq 0$, whence $\Delta(u) \neq 0$. From the elementary theory of the resolution of linear equations, it follows that the basic elements $M_{u,i}$ may be selected in such a way that

a) The elements $\Delta(u)M_{u,i}$ have coördinates which are polynomials in the $u$'s.

b) There exists a system of $l$ indices $\{i_1, i_2, \cdots, i_l\}$ such that the coefficient of $L_{i_\lambda}$ in $M_{u,\rho}$ is $\delta_{\lambda,\rho}$ (i. e., 1 if $\lambda = \rho$, 0 if $\lambda \neq \rho$), $(1 \leqq \lambda, \rho \leqq l)$.

From the condition a), it follows that we can make the substitution $u \rightarrow u_0$ in the coördinates of the elements $M_{u,i}$; we obtain in this way $l$ elements $M_{0,i} \in \mathfrak{L}$. Since we have $(\boldsymbol{L}_u)^l M_{u,i} = 0$, we have $(\boldsymbol{L}_0)^l M_{0,i} = 0$, whence $M_{0,i} \in \mathfrak{N}_0$. From condition b), it follows that the $l$ elements $M_{0,i}$ form a base for $\mathfrak{N}_0$.

Let us now form a system of Plückerian coördinates of $\mathfrak{N}_u$ by means of the base $\{M_{u,1}, \cdots, M_{u,l}\}$. Let $\Xi_\mu(u)$ be the coördinates obtained in this way. It is clear that $\Delta^l(u)\Xi_\mu(u)$ is a polynomial in the quantities $u$, and that the $\Xi_\mu(u_0)$ are the Plückerian coördinates of $\mathfrak{N}_0$ constructed by means of the base $\{M_{0,1}, \cdots, M_{0,l}\}$. Since the quantities $\Xi_\mu(u_0)$ obviously satisfy any homogeneous algebraic equation which is satisfied by the $\Xi_\mu(u)$, they are the coördinates of a point of $[\mathfrak{N}]$: Lemma 2 is proved.

We shall see later that, conversely, the coördinates of a point of $[\mathfrak{N}]$ are the coördinates of some Cartan algebra of $\mathfrak{L}$, provided we exclude the points which lie on a certain number of proper sub-manifolds of $[\mathfrak{N}]$.

LEMMA 3. *The dimension of $[\mathfrak{N}]$ is at most $n - l$.*

Let us introduce $l$ new parameters $v_1, v_2, \cdots, v_l$, algebraically independent of the $u$'s. We set

$$L_{u,v} = \sum_{k=1}^{l} v_k M_{u,k} = \sum_{j=1}^{n} V_j(u, v) L_j$$

where the $V_j$'s are linear forms in $v_1, v_2, \cdots, v_l$ with coefficients in $\Omega(u)$. The element $L_{u,v}$ is clearly a regular element of the Lie algebra $\mathfrak{L}_{u,v} = \Omega(u,v)\mathfrak{L}$; hence it belongs to a Cartan sub-algebra $\mathfrak{N}_{u,v}$ of $\mathfrak{L}_{u,v}$, whose Plückerian coördinates are the quantities $\Xi_\mu(V_1, V_2, \cdots, V_n)$. On the other hand, $L_{u,v}$ is contained in $\Omega(u,v)\mathfrak{N}_u$, which is obviously also a Cartan sub-algebra of $\mathfrak{L}_{u,v}$. Hence the quantities $\Xi_\mu(V)$, $\Xi_\mu(u)$ are proportional.

The field $\Sigma$ of rational functions on $[\mathfrak{N}]$ is the sub-field of $\Omega(u)$ composed of the homogeneous functions of degree 0 of the $\Xi_\mu(u)$. We claim that we can find $l$ of the variables $u$ which are algebraically independent over $\Sigma$. In fact, since the $M_{u,k}$ are linearly independent, we can find $l$ of the forms $V_k$, say $V_{k_1}, V_{k_2}, \cdots, V_{k_l}$ which are linearly independent with respect to the $v$'s. If we had a relation

$$F(u_{k_1}, u_{k_2}, \cdots, u_{k_l}) = 0$$

where $F$ is a polynomial with coefficients in $\Sigma$, we would also have

$$F(V_{k_1}, V_{k_2}, \cdots, V_{k_l}) = 0,$$

since any element of $\Sigma$ is left unchanged by the substitution $u_k \to V_k$. But this is impossible, since the forms $V_{k_1}, \cdots, V_{k_l}$, being linearly independent over $\Omega(u)$, are also algebraically independent over this field.

Since $\Omega(u)$ is of degree of transcendence $n$ over $\Omega$, $\Sigma$ cannot be of a degree of transcendence $> n - l$ over $\Omega$, which proves Lemma 3.

**III. The proof of the theorem.** Let $\mathfrak{N}_0$ be any Cartan algebra of $\mathfrak{L}$. If $S$ is any element of $\mathfrak{N}_0$, the characteristic equation of $S$ has, besides the $l$-uple root 0, $n - l$ other roots $\alpha_1(S), \alpha_2(S), \cdots, \alpha_{n-l}(S)$ (not necessarily distinct). These roots are linear functions of $S$; none of them is equal to 0 if $S$ is regular. If $\alpha(S)$ is a multiple root of order $m$ for every $S \in \mathfrak{N}_0$, we can assign to $\alpha$ a system of $m$ elements $E_{\alpha,1}, \cdots, E_{\alpha,m}$ of $\mathfrak{L}$ such that the following equalities hold for all $S \in \mathfrak{N}_0$:

$$\begin{aligned}
[E_{\alpha,1}, S] &= \alpha(S)E_{\alpha,1}; \ [E_{\alpha,2}, S] = \alpha(S)E_{\alpha,2} + \beta_{1,2}(S)E_{\alpha,1}, \cdots \\
[E_{\alpha,m}, S] &= \alpha(S)E_{\alpha,m} + \beta_{m-1,m}(S)E_{\alpha,m-1} + \cdots + \beta_{1,m}(S)E_{\alpha,1}.
\end{aligned}$$

If we do this for every root $\alpha$, the $n - l$ elements $E_{\alpha,i}$ together with $l$ basic elements of $\mathfrak{N}_0$, constitute a base for $\mathfrak{L}$.[5]

It follows that the base $\{L_1, L_2, \cdots, L_n\}$ may be chosen in such a way that:

---

[5] Cf. *loc. cit.*[1]

1) $L_{n-l+1}, \cdots, L_n$ constitute a base of $\mathfrak{N}_0$;

2) $L_{n-l+1}$ is a regular element of $\mathfrak{N}_0$;

3) If $1 \leqq i \leqq n-l$, each $L_i$ is one of the elements $E_{a,k}$.

4) We have

$$[L_i, S] = \alpha_i(S)L_i + \sum_{j>i} \beta_{ji}(S)L_j$$

for $1 \leqq i \leqq n-l$, $S \in \mathfrak{N}_0$, where $\alpha_i(S)$ is one of the roots $\alpha(S)$. We say that $L_i$ "belongs" to the root $\alpha_i(S)$.

We know also that, if $L_i$ belongs to the root $\alpha_i$, the same holds for every $[L_i, S]$, if $S \in \mathfrak{N}_0$; and that, if $L_i, L_j$ belong to the roots $\alpha_i, \alpha_j$, $[L_i, L_j]$ belongs to the root $\alpha_i(S) + \alpha_j(S)$ if $\alpha_i(S) + \alpha_j(S)$ is a root and is equal to zero otherwise. From this fact and from the fact that there are only a finite number of roots, it follows at once that $\mathbf{L}_i$ is a nilpotent matrix.

Let us introduce $n-l$ new variables $x_1, x_2, \cdots, x_{n-l}$. Since $\mathbf{L}_i$ is nilpotent, the coefficients of the matrix

$$\exp x_i\mathbf{L}_i = \sum_{k=0}^{\infty} (1/k!)x_i^k\mathbf{L}_i^k$$

are polynomials in the $x_i$. We set

$$\sigma(x_1, \cdots, x_{n-l}) = \sigma(x) = (\exp x_1\mathbf{L}_1)(\exp x_2\mathbf{L}_2) \cdots (\exp x_{n-l}\mathbf{L}_{n-l}).$$

The matrix $\sigma(x)$ belongs to the adjoint group of $\mathfrak{G}$ and its coefficients are polynomials in $x_1, x_2, \cdots, x_{n-l}$.

Since $\sigma(x)$ is an operation of the adjoint group, it produces an automorphism of $\mathfrak{L}$, and $\sigma(x)\mathfrak{N}_0$ is a new Cartan algebra $\mathfrak{N}_x$ (we assume that the $x$'s have been adjoined to $\Omega$). A base of $\mathfrak{N}_x$ is composed of the elements $\sigma(x)L_{n-l+h}$ $(1 \leqq h \leqq l)$; let $\xi_\mu(x)$ be the Plückerian coördinates of $\mathfrak{N}_x$ constructed by means of this base. If we give to the quantities $x$ any values in $\Omega$, the $\xi_\mu(x)$ become the coördinates of a point of $[\mathfrak{N}]$. It follows that the polynomials $\xi_\mu(x)$ themselves are the coördinates of a point of $[\mathfrak{N}]$ depending rationally on the parameters $x$.

We shall now prove that the point $(\xi_\mu(x))$ is $(n-l)$-dimensional. In order to prove this, we denote by $\eta_i$ the coördinate $\xi_\mu$ for which $\mu$ is the arrangement composed of the indices $(i, n-l+2, \cdots, n)$ $(1 \leqq i \leqq n-l+1)$. It will be sufficient to prove that the functional determinant of the $n-l$ functions $\dfrac{\eta_i(x)}{\eta_{n-l+1}(x)}$ $(1 \leqq i \leqq n-l)$ *with respect to* $x_1, x_2, \cdots, x_{n-l}$ is $\neq 0$ for $x_1 = x_2 = \cdots = x_{n-l} = 0$.

We denote by $[x]^2$ the ideal of the polynomials in $x_1, x_2, \cdots, x_{n-l}$ generated by the $x_ix_j$ $(1 \leqq i, j \leqq n-l)$. We have

$$\sigma(x)L_{n-l+h} \equiv L_{n-l+h} - \Sigma_i x_i \alpha_i(L_{n-l+h})L_i \\ - \Sigma_{ij} x_i \beta_{ji}(L_{n-l+h})L_j \pmod{[x]^2}$$

where

$$\beta_{ji}(L_{n-h+1}) = 0 \text{ if } j \geqq i.$$

The quantities $\xi_\mu(x)$ are obtained by expanding the external product

$$\sigma(x)L_{n-l+1} \,\square\, \sigma(x)L_{n-l+2} \,\square \cdots \square\, \sigma(x)L_n.$$

It follows that we have

$$\left.\begin{aligned} \eta_{n-l+1}(x) &\equiv 1, \\ \eta_i(x) &\equiv -x_i\alpha_i(L_{n-l+1}) - \sum_{j>i} x_j \beta_{ij}(L_{n-l+1}) \end{aligned}\right\} \pmod{[x]^2}$$

$$(1 \leqq i \leqq n-l).$$

Hence, the value for $x = 0$ of our functional determinant is equal to $(-1)^{n-l}\prod_{i=1}^{n-l} \alpha_i(L_{n-l+1})$: it is different from zero because $L_{n-l+1}$ is regular.

The manifold $[\mathfrak{N}]$ is of dimension $\leqq n-l$ by Lemma 3, and has a point $(\xi_\mu(x))$ of dimension $n-l$. Hence:

LEMMA 4. *The manifold of the Cartan algebras is of dimension $n-l$ and the point $(\xi_\mu(x))$ is a generic point of this manifold.*

We can now prove the following

LEMMA 5. *If we remove from $[\mathfrak{N}]$ the points of a certain system $\mathfrak{M}_0$ of algebraic manifolds of dimensions $< n-l$, the coördinates of any remaining point of $[\mathfrak{N}]$ are the Plückerian coördinates of a Cartan algebra $\mathfrak{N}$ which is the transform of $\mathfrak{N}_0$ by an operation of the form $\sigma(x)$.*

In fact, let us consider the manifold $\mathfrak{X}$ whose generic point has for coördinates $(x_1, x_2, \cdots, x_{n-l}; \xi_\mu(x)/\xi_{\mu_0}(x))$, where $\xi_{\mu_0}(x)$ is the polynomial denoted above by $\eta_{n-l+1}$ (i. e., $\mu_0$ is the arrangement $(n-l+1, \cdots, n)$).

We may consider the $\xi_\mu(x)/\xi_{\mu_0}(x)$ as the coördinates of a point of an affine algebraic manifold $[\mathfrak{N}]^*$. From Lemma 1, it follows that we can find a finite system $\mathfrak{M}^*_0$ of sub-manifolds of $[\mathfrak{N}]^*$ such that every point of $[\mathfrak{N}]^* - \mathfrak{M}^*_0$ is the projection on $[\mathfrak{N}]^*$ of a point of $\mathfrak{X}$. We denote by $\mathfrak{M}_0$ the set of the points $(\bar{\xi})$ of $[\mathfrak{N}]$ which are such that either $\bar{\xi}_{\mu_0} = 0$ or the point of coördinates $\bar{\xi}_\mu(\bar{\xi}_{\mu_0})^{-1}$ lies in $\mathfrak{M}^*_0$. Let $\xi_\mu$ be the coördinates of a point of $[\mathfrak{N}] - \mathfrak{M}_0$; then we have $\xi_{\mu_0} \neq 0$, and there exists a system $(a_1, a_2, \cdots, a_{n-l})$ of elements of $\Omega$ such that

$$\xi_\mu(a)/\xi_{\mu_0}(a) = \xi_\mu/\xi_{\mu_0}.$$

It follows immediately that the $\xi_\mu$ are the Plückerian coördinates of the Cartan algebra $\sigma(a)\mathfrak{N}_0$.

We can now prove our theorem. Let $\mathfrak{N}_1$ be any other Cartan algebra of $\mathfrak{L}$. The manifold $[\mathfrak{N}]$ contains also a finite system $\mathfrak{M}_1$ of manifolds of dimension $< n - l$ such that every point of $[\mathfrak{N}] - \mathfrak{M}_1$ represents a Cartan algebra which may be obtained from $\mathfrak{N}_1$ by an operation of the adjoint group.

We know that these are points on $[\mathfrak{N}]$ which do not belong either to $\mathfrak{M}_0$ or to $\mathfrak{M}_1$. Such a point represents a Cartan algebra which may be obtained from both $\mathfrak{N}_0$ and $\mathfrak{N}_1$ by operations $\sigma_0, \sigma_1$ of the adjoint group. The operation $\sigma_1^{-1}\sigma_0$ transforms $\mathfrak{N}_0$ into $\mathfrak{N}_1$, which proves the theorem.

*Remark.*[6] The elements of the adjoint group which we used are elements of the sub-group generated by those infinitesimal transformations of the adjoint group which are nilpotent. It can be seen without much difficulty that this group is always an *algebraic* distinguished sub-group of the adjoint group. It coincides with the adjoint group in the case where the latter is either semi-simple or nilpotent. In other cases, it may actually be a proper sub-group. However, it always contains the commutator-sub-group of the adjoint group.

PRINCETON UNIVERSITY.

[6] I am indebted to André Weil for this remark, as well as for several others.

# ON THE ERGODIC DYNAMICS OF ALMOST PERIODIC SYSTEMS.*

By NORBERT WIENER and AUREL WINTNER.**

**Introduction.** The classical "trigonometric" developments in celestial mechanics,[1] that is, the developments to which the whole of the second volume of the *Méthodes Nouvelles* is devoted, are purely formal in nature. It was shown by Birkhoff[2] that the *successive stages* of the process defining these expansions are subject on *any finite time range* to certain estimates of the order of magnitude; estimates which, being exactly of the same type as those underlying his theory of non-linear difference equations, can be adapted to local proofs of existence of periodic motions. Correspondingly, the considerations of the astronomers and the investigations of Poincaré and of Birkhoff do not attempt to prove that there exist trigonometric expansions, say

$$F(t) \sim \sum_k \alpha_k e^{i\lambda_k t}, \qquad (\lambda_k \neq \lambda_j \text{ for } k \neq j),$$

which are *Fourier series*, representing for the function $F(t)$ an anharmonic analysis in a sense suggested by Poincaré himself;[3] namely, in the sense that

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-i\mu t} F(t)\, dt = \begin{cases} 0 & \text{if } \mu \neq \lambda_k \text{ for every } k, \\ \alpha_k & \text{if } \mu = \lambda_k \text{ for some } k, \end{cases}$$

where $\mu$ is any real number.

By a recondite connection of Diophantine intricacy, the problem must somehow depend on the celebrated small divisors in celestial mechanics.[4] On the other hand, it turned out that, by virtue of a property of uniformly almost periodic angular variables, the Diophantine small divisors prove to be

* Received March 28, 1941.

** Fellow of the Guggenheim Foundation.

[1] The classical literature of the subject is collected in R. Marcolongo's bibliography, *Il problema dei tre corpi* (Manuali Hoepli, nos. 403-405), 1919, pp. 61-79. For more recent references cf. A. Wintner, *The Analytical Foundations of Celestial Mechanics*, Princeton, 1941, § 523.

[2] For a short presentation, cf. G. D. Birkhoff, *Dynamical Systems*, New York, 1927, Chap. III–Chap. IV, where references are given to the original papers.

[3] H. Poincaré, "Sur la convergence des séries trigonométriques," *Bulletin Astronomique*, vol. 1 (1884), pp. 319-327.

[4] Cf. *loc. cit.*[1], § 523–§ 529.

harmless, not only in the integrable case of Liouville systems,[5] but also in such cases (by no means trivial, though relatively simple) as Hill's theory of the perigee[6] or Adams' theory of the lunar node.[7] This suggested that the treatment of the general problem might be attempted if Bohr's theory of uniformly almost periodic functions, which deals indeed with a situation quite degenerate from the dynamical (or, equivalently, topological) point of view, were replaced by the only result of real generality established for the solutions of dynamical systems; namely, by the ergodic theorem of Birkhoff.[8]

In this direction, it was shown, first in the case of Lagrange's problem of mean motions[9] and then,[10] via the Fourier integral theory of autocorrelations,[11] for any metrically transitive system, that Birkhoff's ergodic theorem actually leads to the existence of an anharmonic analysis, as defined by the above pair of formulae. The case of an arbitrary measure-preserving flow can be reduced to the metrically transitive case, if use is made of von Neumann's transfinite decomposition of the flow into its irreducible components.[12] In this sense, the existence problem for an anharmonic analysis, as formulated above, can be considered as solved in the general case.

However, this general result does not supply the answer to the more specific dynamical question which concerns not merely the existence of an anharmonic analysis but also the problem whether the resulting Fourier series of the function $F(t)$ does or does not converge to $F(t)$ in the mean $(-\infty < t < \infty)$. Obviously, the answer to this specific question of *completeness* cannot be in the affirmative in the general case. For instance, if a surface of constant negative curvature is of finite connectivity and of finite area, then its geodesic flow is, according to Hedlund,[13] a mixture, and so it is easy to see that the anharmonic analysis cannot satisfy the condition of completeness.

---

[5] *Ibid.*, § 194–§ 198. [6] *Ibid.*, § 520–§ 522. [7] *Ibid.*, § 484–§ 487.

[8] G. D. Birkhoff, "Proof of a recurrence theorem for strongly transitive systems," "Proof of the ergodic theorem," *Proceedings of the National Academy of Sciences*, vol. 17 (1931), pp. 650-655, 656-660.

[9] A. Wintner, "On an ergodic analysis of the remainder term of mean motions," *Ibid.*, vol. 26 (1940), pp. 126-128.

[10] N. Wiener and A. Wintner, "Harmonic analysis and ergodic theory," *American Journal of Mathematics*, vol. 63 (1941), pp. 415-426.

[11] N. Wiener, *The Fourier Integral and Certain of its Applications*, Cambridge, 1933, § 20–§ 23.

[12] J. v. Neumann, "Zur Operatorentheorie in der klassischen Mechanik," *Annals of Mathematics*, vol. 33 (1932), pp. 587-642 and 789-791.

[13] G. A. Hedlund, "Fuchsian groups and mixtures," *ibid.*, vol. 40 (1939), pp. 370-383.

The object of the present paper is to delimit the class of those flows on which the condition of completeness is satisfied, i. e., on which the anharmonic analysis (in the sense defined above) becomes a Fourier analysis (in the usual sense of the word).

**1.** In what follows, the existence of an average, $M_t\{g(t)\}$, of a function $g(t)$, $-\infty < t < \infty$, will be meant in the following sense: $g(t)$ is of class $(L)$ on every bounded $t$-interval and there exists a finite limit

$$M\{g(t)\} = \lim_{\substack{-A\to\infty \\ B\to\infty}} \frac{1}{B-A}\int_A^B g(t)\,dt \tag{1}$$

(in other words,

$$\frac{1}{B}\int_0^B g(t)\,dt \quad \text{and} \quad \frac{1}{B}\int_{-B}^0 g(t)\,dt$$

tend to a common finite limit as $B \to \infty$). It will be very important that the limit process occurring in the definition of $M_t\{g(t)\}$ is the one given under (1), and not the limit process

$$B - A \to \infty \tag{1 bis}$$

which is compatible with $B \to -\infty$ or $A \to \infty$. In fact, while Birkhoff's ergodic theorem [8] holds (for almost all points $P$ of the phase space) if its statement is formulated in terms of the average definition (1), it does not,[14] in general, hold (for almost all $P$) if the average is referred to the unrestricted limit process (1 bis). On the other hand, it is well-known that the mean ergodic theorem,[15] resulting by integration with respect to $P$ over the phase space, is sufficiently rough to hold even if the $t$-average involved belongs to the unrestricted limit process (1 bis). The method of proof in § 8 will depend precisely on this discrepancy between the two ergodic theorems.

**2.** Let $f(t)$, $-\infty < t < \infty$, be of class $(L)$ on every bounded $t$-interval. For a fixed real number $x$, the $x$-th Fourier constant of $f$ is defined by

$$a(x) = M_t\{e^{-ixt}f(t)\}, \tag{2}$$

provided that this average exists. If $a(x)$ exists, $a(-x)$ need not exist, since it is not assumed that $\bar{f} = f$. If (2) exists for every $x$, the function $f(t)$

---

[14] Examples to this effect can be constructed from a perusal of the proof of the dominated ergodic theorem of N. Wiener, "The ergodic theorem," *Duke Mathematical Journal*, vol. 5 (1939), pp. 1-18.

[15] Cf. G. Birkhoff, "The mean ergodic theorem," *ibid.*, pp. 19-20.

will be said to possess an amplitude function, $a(x)$. The amplitude function $a(x)$, $-\infty < x < \infty$, if it exists, is a rather discontinuous function in the relevant cases.

Let $f(t)$, $-\infty < t < \infty$, be of class $(L^2)$ on every bounded $t$-interval. If the average

$$c(s) = M_t\{f(t+s)\bar{f}(t)\} \tag{3}$$

exists for every real number $s$, and if $c(s)$ is a continuous function, $f(t)$ is said to possess a correlation function, $c(s)$.

It is known [11] that every function $c(s)$, $-\infty < s < \infty$, which is a correlation function (for a suitable $f$) is the Fourier-Stieltjes transform,

$$c(s) = \int_{-\infty}^{\infty} e^{isx} d\phi(x), \tag{4}$$

of a non-decreasing bounded function $\phi(x)$, $-\infty < x < \infty$, the "periodogram" [16] of $f(t)$. If the monotone function $\phi$ is thought of as normalized, for instance by

$$\text{(4 bis)} \qquad \phi(-\infty) = 0 \quad \text{and} \quad \phi(x-0) = \phi(x), \quad (-\infty < x < \infty),$$

then, according to the uniqueness theorem of the Fourier-Stieltjes transform, the periodogram $\phi$ is uniquely determined by the function (4) and therefore, via (3), by the function $f$ (on the other hand, the function (3) does not, of course, determine the function $f$). It may be mentioned that there exists to every non-decreasing bounded function $\phi$ a function $f$ possessing the given $\phi$ as periodogram.[17]

It is well-known [18] that the Fourier-Stieltjes transform, (4), of any non-decreasing bounded function, $\phi$, has the amplitude function

$$M_s\{e^{-ixs}c(s)\} = \phi(x+0) - \phi(x-0) \tag{5}$$

(so that, in particular,

$$\text{(5 bis)} \qquad M_s\{e^{-ixs}c(s)\} \geqq 0$$

---

[16] This nomenclature, which we now propose in order to eliminate the confusion that both of us had a definition of "spectrum" (cf. *loc. cit.*[10], the fourth and fifth footnotes), differs from Schuster's terminology in the theory of hidden periodicities only insofar as Schuster, having been interested only in cases where $\phi(x)$ is a step function, has called the jump ($\geqq 0$) of $\phi$ at $x$ (that is, the "histogram" of $\phi$), and not $\phi$ itself, the periodogram of $f$.

[17] N. Wiener and A. Wintner, "On singular distributions," *Journal of Mathematics and Physics* (M. I. T.), vol. 17 (1938), pp. 233-246, § 5.

[18] This useful elementary lemma, found by P. Lévy in his *Calcul des probabilités*, Paris, 1925, pp. 169-172, has its historical origin in physical optics.

for $-\infty < x < \infty$). Clearly, (5) remains unchanged if the function $\phi(x)$ occurring in (4) is replaced by its purely discontinuous component [which can be $\equiv 0$]; that is, by the non-decreasing bounded function defined for $-\infty < x < \infty$ by

$$(6) \qquad \sum_{x_n < x} (\phi(x_n + 0) - \phi(x_n - 0)),$$

where $x_1, x_2, \cdots$ is any sequence containing all discontinuity points of $\phi$.

$\phi(x)$ is called a step function if it is identical with the function (6). In this case, and only in this case, the bounded, uniformly continuous function (4), which is almost periodic ($B^2$) even if $\phi(x)$ has a (continuous) singular and/or absolutely continuous component, is a uniformly almost periodic function.

**3.** Let $f(t)$, $-\infty < t < \infty$, be any function which has a correlation function, (3), and is such that the $x$-th Fourier constant, (2), exists for a *fixed* $x$.

Under these assumptions, it is easy to verify that the function of $t$ defined (for the fixed value of $x$) by

$$(7) \qquad f^{(x)}(t) = f(t) - a(x)e^{ixt}$$

possesses a correlation function,

$$(8) \qquad c^{(x)}(s) = M_t\{f^{(x)}(t+s)\bar{f}^{(x)}(t)\},$$

which turns out to be the function

$$(9) \qquad c^{(x)}(s) = c(s) - |a(x)|^2 e^{ixs};$$

and that the $x$-th Fourier constant of this function of $s$ exists and is represented by

$$(10) \qquad M_s\{e^{-ixs}c^{(x)}(s)\} = \phi(x+0) - \phi(x-0) - |a(x)|^2,$$

where $\phi$ denotes the periodogram of $f(t)$.

In fact, since (3) exists for every $s$, substitution of (7) into (8) gives

$$c^{(x)}(s) = c(s) - a(x)e^{ixs}M_t\{\bar{f}(t)e^{ixt}\} - \bar{a}(x)M_t\{f(t+s)e^{-ixt}\} + |a(x)|^2 e^{ixs},$$

provided that the averages

$$M_t\{\bar{f}(t)e^{ixt}\} \quad \text{and} \quad M_t\{f(t+s)e^{-ixt}\} \equiv M_t\{f(t)e^{-ixt}e^{ixs}\}$$

exist. But they exist, and are represented by $\bar{a}(x)$ and $a(x)e^{isx}$ respectively, since (2) is supposed to exist for the given value of $x$. Accordingly,

$$c^{(x)}(s) = c(s) - a(x)e^{ixs}\bar{a}(x) - \bar{a}(x)a(x)e^{isx} + |a(x)|^2 e^{ixs}.$$

Since this reduces to (9), and since (9) and (5) imply (10), the proof is complete.

Its trivial character notwithstanding, the result of this simple calculation will be fundamental in the sequel, for the following reason: (10) *can be interpreted as a commutation rule, expressing the deviation which results by subjecting the function* $e^{-ixs}\bar{f}(t)f(t+s)$ *of the two variables* $s, t$ (*where* $x$ *is fixed*) *to the two iterated average operators* $M_tM_s, M_sM_t$; *the error committed by the interchange of the two limit processes* $M_t, M_s$ *having precisely the value* (10).

In fact, from (2), where $x$ is fixed,

$$|a(x)|^2 = M_t\{e^{ixt}\bar{f}(t)\}M_s\{e^{-ixs}f(s)\} \equiv M_t\{e^{ixt}\bar{f}(t)M_s\{e^{-ixs}f(s)\}\};$$

so that, since $M_s\{e^{-ixs}f(s)\} = e^{-itx}M_s\{e^{-ixs}f(t+s)\}$ for every $t$,

$$|a(x)|^2 = M_t\{\bar{f}(t)M_s\{e^{-ixs}f(t+s)\}\} \equiv M_t\{M_s\{e^{-ixs}f(t+s)\bar{f}(t)\}\}.$$

On the other hand, substitution of (3) into (5) gives

$$\begin{aligned}\phi(x+0)-\phi(x-0) &= M_s\{e^{-ixs}M_t\{f(t+s)\bar{f}(t)\}\}\\ &\equiv M_s\{M_t\{e^{-ixs}f(t+s)\bar{f}(t)\}\}.\end{aligned}$$

Hence, (10) can be written as a commutator,

(10 bis) $$M_s\{e^{-ixs}c^{(x)}(s)\} = [M_sM_t - M_tM_s]e^{-ixs}f(t+s)\bar{f}(t).$$

In particular

$(11_1)$ $$M_s\{M_t\{e^{-ixs}f(t+s)\bar{f}(t)\}\} = M_t\{M_s\{e^{-ixs}f(t+s)\bar{f}(t)\}\}$$

holds, for the given value of $x$, if and only if

$(11_2)$ $$M_s\{e^{-ixs}c^{(x)}(s)\} = 0$$

for the same $x$.

The above results can be restated as follows:

THEOREM 1. *If a function* $f(t)$, $-\infty < t < \infty$, *has a correlation function*, (3), *and if the* $x$*-th Fourier constant*, (2), *of* $f(t)$ *exists for a fixed* $x$, *then the periodogram*, $\phi$, *of* $f(t)$ *is subject, at the given point* $x$, *to the inequality*

(12) $$|a(x)|^2 \leqq \phi(x+0)-\phi(x-0),$$

*where the sign of equality holds if and only if the commutability condition* $(11_1)$ *is satisfied for the given* $x$.

In fact, since (7) has a correlation function, (8), it has a periodogram, say $\phi^{(x)} = \phi^{(x)}(y)$, $-\infty < y < \infty$; so that, corresponding to (4),

$$c^{(x)}(s) = \int_{-\infty}^{\infty} e^{iys} d\phi^{(x)}(y).$$

On applying the corollary (5 bis) of (5) to this Fourier-Stieltjes transform instead of to (4), one obtains the inequality

$$M_s\{e^{-iys}c^{(x)}(s)\} \geqq 0$$

for $-\infty < y < \infty$. The particular case $y = x$ of this inequality, when combined with the identity (10), implies Theorem 1, since $(11_2)$ has been seen to be equivalent to $(11_1)$.

**4.** The proof of Theorem 3 (concerning almost periodic ($B^2$) functions; cf. § 5 below) will depend (§ 6) on Theorem 1 alone. On the other hand, before Theorem 3 becomes applicable to the ergodic problem, it will (in § 9) be necessary to use, besides Theorem 3, a Tauberian counterpart of Theorem 1, namely the following result:[19]

THEOREM 2. *If a bounded function $f(t)$, $-\infty < t < \infty$, has a correlation function,* (3)–(4), *then the x-th Fourier constant,* (2), *of $f(t)$ exists and vanishes for all those values $x$ which are continuity points of the periodogram, $\phi(x)$, of $f(t)$.*

The Tauberian element is represented by the assumption $|f(t)| <$ const. of Theorem 2.

*Loc. cit.*[10], the proof of Theorem 2 was given for sequences, instead of for functions, but it was pointed out there that the proof applies without change to the case of functions as well. It is clear also from the proof that the assumption $|f(t)| <$ const. of Theorem 2 can be replaced by other, more general, Tauberian conditions; for instance, by the restriction $f(t) \geqq 0$, if $f(t)$ is real-valued. However, it is undecided whether or not the assertion of Theorem 2 is true for *every* function possessing a correlation function.

**5.** Let the almost periodicity ($B^2$) of a function $f(t)$, $-\infty < t < \infty$, be meant in the sense that the trigonometric approximability in quadratic mean refers to the definition (1) of the average.

It will be shown that, in terms of the notions considered in § 2 and § 3, the almost periodicity ($B^2$) of a function $f(t)$ can be characterized, without

[19] N. Wiener and A. Wintner, *loc. cit.*[10], Lemma.

any explicit reference to the notions of approximation, translation or compactness, as follows:

THEOREM 3. *A function* $f(t)$, $-\infty < t < \infty$, *is almost periodic* $(B^2)$ *if and only if it has*

(I) *an amplitude function;*
(II) *a correlation function;*
(III) *a periodogram which is a step function;*
(IV) *the commutability property* $(11_1)$ *for every* $x$.

As an illustration, let $f(t) = \sin | t^{\frac{1}{2}} |$. Then it is easily verified that $M_t\{| f(t) |^2\} \neq 0$, and that the functions (2), (3) exist and reduce to constants: $a(x) \equiv 0$, $c(s) \equiv M_t\{| f(t) |^2\}$. Since $c(s) = \text{const.}$ means, by (4), that $\phi(x)$ is a constant multiple of the step function $\operatorname{sgn} x$, it follows that (I), (II), (III) are satisfied. But $f(t)$ is not almost periodic $(B^2)$, since $M_t\{| f(t) |^2\} \neq 0$ and the Parseval relation are at variance with $a(x) \equiv 0$. Accordingly, the conditions (I), (II), (III) together are not sufficient for almost periodicity $(B^2)$; that they are necessary, is quite straightforward (cf. the beginning of § 6 below). Thus the emphasis in Theorem 2 is on the form of the additional condition, (IV).

That (IV) becomes, in virtue of (I), (II), (III), equivalent to the definition of almost periodicity $(B^2)$, is not as surprising as appears at first glance. In fact, the validity of $(11_1)$ for every $x$ is equivalent to the assumption that $(11_2)$ is an identity in $x$. But $(11_2)$ can be thought of as an orthogonality relation (for every fixed $x$). Then the criterion represented by the set of the four conditions (I)–(IV) corresponds to the definition of the almost periodic class $(B^2)$ (that is, to the condition of trigonometric approximability in quadratic mean) in the same way that the notion of a *closed* sequence of orthogonal functions relates to the equivalent notion of a *complete* sequence of orthogonal functions in the Hilbert space of the functions $f(t)$, $0 \leqq t \leqq 1$, of class $(L^2)$. Of course, the situation is more delicate in the present case, since the integrals become replaced by averages and, correspondingly, the index $x$ cannot be restricted to an enumerable set which is independent of $f$.

**6.** In order to prove Theorem 3, suppose first that $f(t)$ is almost periodic $(B^2)$. Then (2) exists for every $x$, and vanishes except when $x$ belongs to a set which is enumerable (at most). If $x_1, x_2, \cdots$ is a sequence of distinct numbers containing this set, then, by the Parseval relation,

$$(13) \qquad \sum_n | a(x_n) |^2 < \infty.$$

In these notations, the Fourier expansion $(B^2)$ of $f(t)$ is

$$f(t) \sim \sum_n a(x_n) e^{ix_n t}.$$

Furthermore, if $s$ is arbitrarily fixed, the translated function, $f(t+s)$, of $t$ is almost periodic $(B^2)$, and

$$f(t+s) \sim \sum_n a(x_n) e^{ix_n s} e^{ix_n t}.$$

Hence, by the polarized form of the Parseval relation,

$$M_t\{f(t+s)\bar{f}(t)\} = \sum_n | a(x_n) |^2 e^{ix_n s}.$$

This function of $s$ is, by (12), continuous (and, as a matter of fact, uniformly almost periodic), and appears in the Fourier-Stieltjes form (4) if one puts

$$(14) \qquad \phi(x) = \sum_{x_n < x} | a(x_n) |^2;$$

it being assured by (13) that (14) defines a non-decreasing bounded function for $-\infty < x < \infty$. Furthermore, (14) can be written in the form (6), where

$$(15) \qquad \phi(x_n + 0) - \phi(x_n - 0) = | a(x_n) |^2;$$

so that $\phi(x)$ is purely discontinuous.

This completes the proof of the necessity of the conditions (I), (II), (III), if $f(t)$ is almost periodic $(B^2)$.

Next, suppose that a given $f(t)$, which need not be almost periodic $(B^2)$, satisfies the three conditions (I), (II), (III). The two conditions (I), (II) together can be expressed by saying that the assumptions of Theorem 1 are satisfied not for a fixed $x$ but for every $x$; while condition (III) means that there exists a sequence of distinct numbers $x_1, x_2, \cdots$ by means of which the non-decreasing bounded function $\phi(x)$, $-\infty < x < \infty$, implicitly defined by (3), (4) and (4 bis), can be represented in the form (6).

Under these assumptions, define, for every positive integer $N$, a function $f_N(t)$, $-\infty < t < \infty$, by placing

$$(16) \qquad f_N(t) = f(t) - \sum_{n \leq N} a(x_n) e^{ix_n t}.$$

Since (7) has a correlation function (8) which reduces to (9), it follows from the orthogonality relation

$$(17) \qquad M_t\{e^{ixt} e^{-iyt}\} = \begin{cases} 0, & x \neq y \\ 1, & x = y \end{cases}$$

by induction from $N-1$ to $N$, that (16) has a correlation function

(18) $$c_N(s) = M_t\{f_N(t+s)\bar{f}_N(t)\}$$

which reduces to

(19) $$c_N(s) = c(s) - \sum_{n \leqq N} |a(x_n)|^2 e^{ix_n s}.$$

Since the assumptions of Theorem 1 are satisfied for every $x$, the inequality (12) holds for $-\infty < x < \infty$; so that

(20) $$|a(x_n)|^2 \leqq \phi(x_n + 0) - \phi(x_n - 0)$$

for every $n$. This implies (13), since

(21) $$\sum_n (\phi(x_n + 0) - \phi(x_n - 0)) < \infty,$$

$\phi(x)$ being a non-decreasing bounded function for $-\infty < x < \infty$. Furthermore, since $x_n \neq x_m$ for $n \neq m$, it is clear from (20) and (21) that the function

(22) $$\phi_N(x) = \sum_{\substack{x_n < x \\ n \leq N}} (\phi(x_n + 0) - \phi(x_n - 0) - |a(x_n)|^2) + \sum_{\substack{x_n < x \\ n > N}} (\phi(x_n + 0) - \phi(x_n - 0))$$

is non-decreasing and bounded for $-\infty < x < \infty$. Since $\phi(x)$ is supposed to be the sum (6), one sees from (22), (19) and (4) that

(23) $$c_N(s) = \int_{-\infty}^{\infty} e^{isx} d\phi_N(x);$$

so that, since (18) is the correlation function of (16), the function (22) is the periodogram of (16). In particular

(24) $$M_t\{|f_N(t)|^2\} = \phi_N(\infty) - \phi_N(-\infty),$$

as is seen by placing $s = 0$ in (18) and (23).

All of this was deduced under the assumption that $f(t)$ is a function which satisfies (I), (II), (III). In order to complete the proof of Theorem 3, it remains to be shown that, if (I), (II), (III) are satisfied, $f(t)$ is or is not almost periodic ($B^2$) according as it does or does not satisfy (IV).

By definition, $f(t)$ is almost periodic ($B^2$) if and only if the functions (16) satisfy the requirement

$$M_t\{|f_N(t)|^2\} \to 0, \qquad (N \to \infty).$$

According to (24), this can be written in the form

$$\phi_N(\infty) - \phi_N(-\infty) \to 0, \qquad (N \to \infty).$$

This condition is reduced by (20), (21) and (22) to the equivalent requirement

$$\sum_n (\phi(x_n + 0) - \phi(x_n - 0) - |a(x_n)|^2)^2 = 0.$$

According to (20), this requirement is satisfied if and only if (15) holds for every $n$. Since it is not assumed that the Fourier constant $a(x_n)$ be distinct from 0 for $n = 1, 2, \cdots$, one can adjoin every given real number $x$ to the sequence $x_1, x_2, \cdots$. In this sense, the requirement that (15) holds for every $n$ is equivalent to the condition that the function (10) of $x$ should vanish identically. But this is precisely the condition (IV), since ($11_1$) is equivalent to ($11_2$).

This completes the proof of Theorem 3.

7. The results of § 3–§ 6, which refer to a *single* function $f(t)$, will now be applied to the set of almost all functions in Birkhoff's ergodic theorem. To this end, it will first be necessary to collect those straightforward consequences of this theorem which involve the notions discussed in § 2.

Let a space $S$ of points $P$ carry a Lebesgue measure for which $S$ is of finite measure, and let $T^t$, $-\infty < t < \infty$, be a cyclic group of transformations which map $S$ on itself, are measure-preserving, and such that the functions $T^tP$ of $(t, P)$ satisfies the usual measurability condition. Then Birkhoff's ergodic theorem states that, for every function $f(P)$ of class $(L)$ on $S$, the average $M_t\{f(T^tP)\}$ exists for almost all $P$, and represents a function $f^*(P)$ of class $(L)$ on $S$.

Let $R$ be a *fixed* rotation of a circle $C$ on which the arc length is assigned as Lebesgue measure. Consider on the product space $S \times C$ the product flow $T^t \times R$ and the corresponding product measure, and apply Birkhoff's ergodic theorem to this product model. It is then clear from Fubini's theorem that, if $x$ is any *fixed* real number, the $x$-th Fourier constant,

$$a_P{}^f(x) = M_t\{e^{-ixt}f(T^tP)\}, \tag{25}$$

of $f(t) = f(T^tP)$ exists, for every given $f(P)$ of class $(L)$, for almost all $P$.

The proof of this well-known trivial consequence of Birkhoff's ergodic theorem has been given here only because its analogue corresponding to the mean ergodic theorem will also be needed. Let $f(P)$ be of class $(L^2)$ on $S$. It has been pointed out by Wiener [20] that application of Birkhoff's ergodic theorem to a product model, when combined, as before, with Fubini's theorem, assures the existence of the correlation function

$$(26) \qquad c_P{}^f(s) = M_t\{f(T^{t+s}P)\bar{f}(T^tP)\}$$

of $f(t) = f(T^tP)$ for almost all $P$. The corresponding fact supplied by the mean ergodic theorem (in the mean of the function space $(L) = (L^1)$ on $S$) is that

$$(27) \qquad \lim_{B-A\to\infty} \int_S \left| c_P{}^f(s) - \frac{1}{B-A}\int_A^B f(T^{t+s}P)\bar{f}(T^tP)\,dt \right| d_PS = 0$$

holds for $-\infty < s < \infty$ (it is understood that $d_PS$ denotes the volume element on $S$).

Let

$$(28) \qquad \phi_P{}^f = \phi_P{}^f(x), \qquad -\infty < x < \infty,$$

denote the periodogram of $f(t) = f(T^tP)$, $-\infty < t < \infty$, (for almost all $P$); so that, corresponding to (4),

$$(29) \qquad c_P{}^f(s) = \int_{-\infty}^{\infty} e^{isx} d_x\phi_P{}^f(x), \qquad -\infty < s < \infty,$$

and, by (5)

$$(30) \qquad M_s\{e^{-ixs}c_P{}^f(s)\} = \phi_P{}^f(x+0) - \phi_P{}^f(x-0) \text{ for every } x.$$

It turns out that, if $x$ is any *fixed* real number, then the averages (25) and (26), which exist for almost all $P$, satisfy the relation

$$(31) \qquad M_s\{e^{-ixs}c_P{}^f(s)\} = |\, a_P{}^f(x)\,|^2$$

for almost all $P$. It is precisely (31) that will make applicable Theorems 1 and 3. Corresponding to this central rôle of (31), the proof of (31) is somewhat lengthy; it will occupy the whole of § 8.

---

[20] Cf. *loc. cit.*[10], the eighth footnote.

**8.** Let $f(P)$ be of class $(L^2)$ on $S$. According to the mean ergodic theorem, there exists on $S$ a function $f^*(P)$ of class $(L^2)$ such that, *in the sense of* (1 bis),

$$\lim_{B-A\to\infty} \int\limits_S | f^*(P) - \frac{1}{B-A} \int\limits_A^B f(T^tP)dt\,|^2\, d_PS = 0.$$

Hence, if $x$ is any real number, it follows, by a repetition of the product argument which led from the Birkhoff ergodic theorem to (25), that there exists for the given $x$ a function $f^*_x(P)$ of class $(L^2)$ on $S$ such that

$$\lim_{B-A\to\infty} \int\limits_S | f^*_x(P) - \frac{1}{B-A} \int\limits_A^B e^{-ixt} f(T^tP)dt\,|^2\, d_PS = 0.$$

Since this means convergence in the mean $(L^2)$ of $S$, it implies the existence of a subsequence which is convergent almost everywhere on $S$. Thus there exist two sequences, say $A_1, A_2, \cdots$ and $B_1, B_2, \cdots$, such that

$$A_n \to -\infty \quad \text{and} \quad B_n \to \infty \quad \text{as} \quad n \to \infty$$

and, for almost all $P$,

$$\frac{1}{B_n - A_n} \int\limits_{A_n}^{B_n} e^{-ixs} f(T^tP)dt \to f^*_x(P) \quad \text{as} \quad n \to \infty,$$

where $x$ is fixed. Since (25) exists for almost all $P$, it follows that

$$f^*_x(P) = a_P{}^f(x).$$

On substituting this into the definition of $f^*_x(P)$, one sees that

$$\lim_{B-A\to\infty} \int\limits_S | a_P{}^f(x) - \frac{1}{B-A} \int\limits_A^B e^{-ixt} f(T^tP)dt\,|^2\, d_PS = 0.$$

The complex conjugate of this (real) relation is

$$\lim_{D-C\to\infty} \int\limits_S | \bar{a}_P{}^f(x) - \frac{1}{D-C} \int\limits_C^D e^{ixs} \bar{f}(T^sP)ds\,|^2\, d_PS = 0,$$

where $s, C, D$ are written in place of $t, A, B$. Since the quadratic mean on $S$ has the properties of a distance (Schwarz, Minkowski), the last two relations imply that

$$\lim_{\substack{B-A\to\infty \\ D-C\to\infty}} \int_S \left| \, |a_P{}^f(x)|^2 - \frac{1}{B-A}\int_A^B e^{-ixt} f(T^tP)\,dt \, \frac{1}{D-C}\int_C^D e^{ixs}\bar{f}(T^sP)\,ds \, \right| d_PS = 0,$$

where the limit sign refers to a *double* limit, $(A, B)$ and $(C, D)$ being independent of one another.

Since the time ranges, $A \leqq t \leqq B$ and $C \leqq s \leqq D$, are chosen independently, and since the length $D - C$ remains unchanged if the range $C \leqq s \leqq D$ is replaced by $C + t \leqq s \leqq D + t$, where $t$ is arbitrary, the last relation can be written in the form

$$\lim_{\substack{B-A\to\infty \\ D-C\to\infty}} \int_S \left| \, |a_P{}^f(x)|^2 - \frac{1}{B-A}\,\frac{1}{D-C}\int_A^B e^{-ixt} f(T^tP)\,dt \int_{C+t}^{D+t} e^{ixs}\bar{f}(T^sP)\,ds \, \right| d_PS = 0.$$

But

$$\int_{C+t}^{D+t} e^{ixs}\bar{f}(T^sP)\,ds = e^{ixt}\int_C^D \bar{f}(T^{t+s}P)\,ds;$$

so that, by Fubini's theorem,

$$\int_A^B e^{-ixt} f(T^tP)\,dt \int_{C+t}^{D+t} e^{ixs}\bar{f}(T^sP)\,ds = \int_C^D e^{ixs}ds \int_A^B \bar{f}(T^{t+s}P) f(T^tP)\,dt.$$

Hence, the complex conjugate of the preceding limit relation can be written in the form

$$\lim_{\substack{B-A\to\infty \\ D-C\to\infty}} \int_S \left| \, |a_P{}^f(x)|^2 - \frac{1}{B-A}\,\frac{1}{D-C}\int_C^D e^{-ixs}\Big[\int_A^B f(T^{t+s}P)\bar{f}(T^tP)\,dt\Big]ds \, \right| d_PS = 0.$$

Since this relation expresses convergence in the mean of $(L) = (L^1)$ on $S$, it implies the existence of a suitable subsequence which is convergent almost everywhere on $S$. Thus there exist, for the fixed value of $x$, two pairs of sequences of numbers, say

$$A_1, A_2, \cdots, B_1, B_2, \cdots \quad \text{and} \quad C_1, C_2, \cdots, D_1, D_2, \cdots,$$

such that

$$A_n \to -\infty,\ B_n \to \infty \text{ as } n \to \infty \text{ and } C_m \to -\infty,\ D_m \to \infty \text{ as } m \to \infty$$

and, for almost all $P$,

$$\lim_{\substack{n\to\infty\\ m\to\infty}} \frac{1}{B_n - A_n} \frac{1}{D_m - C_m} \int_{C_m}^{D_m} e^{-ixs}\Big[\int_{A_n}^{B_n} f(T^{t+s}P)\bar{f}(T^tP)\,dt\Big]ds = |\, a_P{}^f(x)\,|^2.$$

Since the limit on the left is a double limit, it can be written as an iterated limit; so that

$$\lim_{m\to\infty} \frac{1}{D_m - C_m} \lim_{n\to\infty} \int_{C_m}^{D_m} \frac{e^{-ixs}}{B_n - A_n} \int_{A_n}^{B_n} f(T^{t+s}P)\bar{f}(T^tP)\,dt\,ds = |\, a_P{}^f(x)\,|^2$$

for almost all $P$.

Since $f(P)$ is of class $(L^2)$ on $S$, there exists, by Birkhoff's ergodic theorem, a finite limit

$$\lim_{\substack{-A\to\infty\\ B\to\infty}} \frac{1}{B-A} \int_A^B |\, f(T^tP)\,|^2\,dt = M_t\{|\, f(T^tP)\,|^2\}$$

for almost all $P$. Hence, it is easily seen from the integrability properties of $f$ and $T^tP$, that, if two numbers, say $C$ and $D(> C)$, are arbitrarily fixed, and if the point $P$ is fixed in the complement of a set of measure zero, then the function

$$\frac{1}{B-A} \int_A^B |\, f(T^{t+s}P)\,|^2\,dt$$

of $s$ remains for $C \leqq s \leqq D$ under a fixed bound [21] (depending on $C$, $D$, and $P$), as $A \to -\infty$, $B \to \infty$. Thus it is clear from the Schwarz inequality and from the assumption

$$A_n \to -\infty, \quad B_n \to \infty, \qquad (n \to \infty),$$

that, if $m$ and $P$ are fixed, then, unless $P$ belongs to a set of measure zero, the functions

$$\frac{e^{-ixs}}{B_n - A_n} \int_{A_n}^{B_n} f(T^{t+s}P)\bar{f}(T^t)\,dt$$

of $s$, where $n = 1, 2, \cdots$, are uniformly bounded on the interval $C_m \leqq s \leqq D_m$. Furthermore, since the average (26), defined by (1), exists for $-\infty < s < \infty$ unless $P$ belongs to a set of measure zero, the expression in the last formula

[21] For the details of the rather elementary proof, cf. N. Wiener, *loc. cit.*[11], p. 155.

line tends for $C_m \leqq s \leqq D_m$ and for almost all $P$ to the limit $c_P{}^f(s)$, as $n \to \infty$. It follows, therefore, from Lebesgue's theorem on term-by-term integration, that

$$\lim_{n\to\infty} \int_{C_m}^{D_m} \frac{e^{-ixs}}{B_n - A_n} \int_{A_n}^{B_n} f(T^{t+s}P)\bar{f}(T^tP)\,dt\,ds = \int_{C_m}^{D_m} e^{-ixs} c_P{}^f(s)\,ds$$

for almost all $P$ and for every $m$.

Accordingly, the relation containing the iterated limit reduces to

$$\lim_{m\to\infty} \frac{1}{D_m - C_m} \int_{C_m}^{D_m} e^{-ixs} c_P{}^f(s)\,ds = |\, a_P{}^f(x)\,|^2$$

for almost all $P$. But the average on the left of (30) exists whenever the correlation function (29) exists; so that, since the latter exists, by § 7, for almost all $P$,

$$\lim_{m\to\infty} \frac{1}{D_m - C_m} \int_{C_m}^{D_m} e^{-ixs} c_P{}^f(s)\,ds = \lim_{\substack{-C\to\infty \\ D\to\infty}} \frac{1}{D - C} \int_{C}^{D} e^{-ixs} c_P{}^f(s)\,ds \equiv M_s\{e^{-ixs} c_P{}^f(s)\}$$

holds, in view of $\lim_{m\to\infty} C_m = -\infty$, $\lim_{m\to\infty} D_m = \infty$, for almost all $P$.

The last two formula lines show that the proof of (31) is now complete; it being understood that a zero set of points $P$ is excluded in (31) for every fixed $x$.

**9.** It is now easy to prove a theorem which in its simplest case may be formulated as follows:

*If $f(P)$ is a bounded, measurable function on a space $S$ of finite measure, and if the flow $T^t$ is metrically transitive on $S$, then the function $f(T^tP)$, $-\infty < t < \infty$, is or is not almost periodic ($B^2$) for almost all $P$ according as its periodogram (which, by § 7, exists for almost all $P$) is or is not a step function for almost all $P$.*

In fact, the metrical transitivity of $T^t$ means that $M_t\{g(T^tP)\}$ is constant almost everywhere on $S$ for every function $g(P)$ of class $(L)$. But then it is clear from the proof of the existence of the correlation functions (26), that (26) is independent of $P$ for almost all $P$. It follows, therefore, from (29) and from the uniqueness theorem of the Fourier-Stieltjes transform,

that the monotone function (28) of $x$ is independent of $P$ for almost all $P$. Hence, the preceding italicized statement is implied by the following theorem:

THEOREM 4. *Let $f(P)$ be a bounded, measurable function on a space $S$ of finite measure, and let $T^tP$, $-\infty < t \infty$, be a measure-preserving, $(t, P)$-measurable transformation group of $S$ into itself; so that, by § 7, the function $f(T^tP)$, $-\infty < t < \infty$, has a periodogram $\phi_P{}^f(x)$, $-\infty < x < \infty$, for almost all $P$. Suppose that there exists an enumerable set of points $x$ which contains the set $X_P{}^f$ for almost all $P$, where $X_P{}^f$ denotes, for almost all $P$, the sequence of those points $x$ at which the monotone function $\phi_P{}^f$ is discontinuous. Then the function $f(T^tP)$, $-\infty < t < \infty$, is almost periodic $(B^2)$ for almost all $P$ if and only if the function $\phi_P{}^f(x)$ of $x$ is a step function of $x$ for almost all $P$.*

In the proof of Theorem 4, the assumption that $f(P)$ is a bounded function will be used only via Theorem 2. Hence, if the answer to the question formulated at the end of § 4 were in the affirmative, Theorem 4 would follow for every function $f(P)$ of class $(L^2)$. Incidentally, it will follow from Theorem 6 below, that every function of class $(L^2)$ can be admitted in the particular case italicized before Theorem 4.

In order to prove Theorem 4, let $f(P)$ be a fixed, bounded, measurable function on $S$. According to § 7, there exists, for every real number $x$, a zero set, say $Z_x$, such that the $x$-th Fourier constant, $a_P{}^f(x)$, of $f(T^tP)$ exists for every point $P$ of $S - Z_x$. Since $f(T^tP)$ has, by § 7, a correlation function, $c_P{}^f$, and therefore a periodogram, $\phi_P{}^f$, for almost all $P$, the zero set $Z_x$ can so be chosen that not only the $x$-th Fourier constant but also the periodogram of $f(T^tP)$ exists for every point $P$ of $S - Z_x$. Finally, since (31) holds for almost all $P$ for every fixed $x$, the zero set $Z_x$ can so be chosen that (31) is satisfied by every point $P$ of $S - Z_x$.

According to the hypothesis of Theorem 4, there exists a zero set, say $Z^*$, and a sequence, say $x_1, x_2, \cdots$, such that, if $P$ is any point of $S - Z^*$, then the monotone function $\phi_P{}^f(x)$, $-\infty < x < \infty$, is defined (i. e., $f(T^tP)$ has a correlation function) and has no discontinuity points $x$ distinct from every $x_n$. It follows, therefore, from Theorem 2, that, if $P$ is any point of $S - Z^*$, the $x$-th Fourier constant, $a_P{}^f(x)$, of $f(T^tP)$ exists and vanishes for every $x$ distinct from every $x_n$. Thus, if $P$ is any point of $S - Z^*$, then

$$\phi_P{}^f(x+0) - \phi_P{}^f(x-0) = 0 = | a_P{}^f(x) |^2$$

for every $x$ distinct from every $x_n$. Hence, it is clear from (30), that, if $P$ is any point of $S - Z^*$, then (31) is satisfied for every $x$ distinct from every $x_n$.

This, when compared with the above definition of the zero sets $Z_x$, $-\infty < x < \infty$, implies that, if the point $P$ is in none of the zero sets

$$Z^*;\ Z_{x_1}, Z_{x_2}, \cdots,$$

then the $x$-th Fourier constant of $f(T^tP)$ exists for every $x$, the correlation function of $f(T^tP)$ exists, and (31) is satisfied for every $x$. In view of (30) and of the equivalence of the two relations $(11_1)$, $(11_2)$, one can express this by saying that conditions (I), (II), (IV) of Theorem 3 are satisfied by $f(t) = f(T^tP)$ for every point $P$ not contained in the zero set

$$Z^* + Z_{x_1} + Z_{x_2} + \cdots.$$

Since Theorem 3 implies that, in case (I), (II), (IV) are satisfied, (III) is necessary and sufficient for the almost periodicity $(B^2)$ of $f(t)$, the proof of Theorem 4 is complete.

It is clear from this proof that the assertion of Theorem 4 supplying a *necessary* condition for the almost periodicity $(B^2)$ of almost all function $f(T^tP)$ certainly holds for arbitrary, and not only for bounded, functions $f(P)$ of class $(L^2)$, and is, in addition, independent of the hypothesis of an enumerable set $x_1, x_2, \cdots$.

**10.** The assumptions of Theorem 4 are so specific that if its assertion is true for *certain* bounded, measurable functions $f(P)$ on $S$ and for a given flow, its assertion need not be true for *every* bounded, measurable function $f(P)$ on $S$ and for the same flow. Thus there arises the question as to a criterion which corresponds to that supplied by Theorem 4 but involves only the flow on $S$, without involving a particular function $f(P)$ on $S$. In order to deduce from Theorem 4 such a criterion, *it will from now on be assumed that there exists on $S$ a complete orthogonal system.* This assumption is made because it will be necessary to establish the connection between the periodograms $\phi_P{}^f$, belonging to individual points $P$ of $S$, on the one hand, and the considerations of Carleman [22] and Koopman,[23] concerning integrals of $f(T^tP)$ over $S$, on the other hand.

---

[22] T. Carleman, "Application de la théorie des équations intégrales linéaires aux équations différentielles de la dynamique," *Arkiv för Mat., Astr. och Fys.*, vol. 22 (1931), no. 7.

[23] B. O. Koopman, "Hamiltonian systems and linear transformations in Hilbert space," *Proceedings of the National Academy of Sciences*, vol. 17 (1931), pp. 315-318.

The latter considerations center around the remark that, if the transition from a function $f_0 \equiv f(P)$ of class $(L^2)$ on $S$ to the function $f_t \equiv f(T^t P)$, where $t$ is fixed, is thought of as a transformation, $f_t = U^t f_0$, of the Hilbert space $(L^2)$ of $S$ into itself, then the measure-preserving property of $T^t$ on $S$ can be expressed by saying that $U^t$ is unitary.[24] Correspondingly, the group property of $T^t$ and the $(P, t)$-measurability of $T^t P$ mean that $U^t$, $-\infty < t < \infty$, is a cyclic group such that, in the sense of strong convergence, $U^s \rightarrow U^t$ whenever $s \rightarrow t$.

Let $f.g$ denote the scalar product,

$$(32) \qquad f.g = \int_S f(P)\bar{g}(P)\,d_P S,$$

of two vectors of the Hilbert space $(L^2)$ on $S$, and let

$$(33) \qquad \bar{f}.\Phi(x)f, \qquad -\infty < x < \infty,$$

be the spectral form of the group of unitary forms,

$$(34) \qquad \bar{f}.U^t f, \qquad -\infty < t < \infty.$$

Thus, if $x$ is fixed, (33) is a bounded Hermitian form in $f$ (boundedness being meant in the sense of Hilbert); while if $f$ is fixed, (33) is a non-negative, non-decreasing, bounded function of $x$ and can be chosen as normalized, corresponding to (4 bis), by

$$\bar{f}.\Phi(-\infty)f = 0 \text{ and } \bar{f}.\Phi(x-0)f = \bar{f}.\Phi(x)f; \text{ so that } \bar{f}.\Phi(\infty)f = \bar{f}.f.$$

Then $U^t$, $-\infty < t < \infty$, determines, according to Stone,[25] exactly one $\Phi(x)$, $-\infty < x < \infty$, such that (34) is representable by means of (33) in the form

---

[24] Since every unitary operator is bounded, and since every bounded operator in Hilbert space is, according to Toeplitz, a bounded matrix, the spectral theory of unitary matrices (A. Wintner, *loc. cit.*[25]) has *not* been generalized by replacing it by the corresponding theory of unitary operators. In this connection, cf. A. Wintner, "Dynamische Systeme und unitäre Matrizen," *Mathematische Zeitschrift*, vol. 36 (1933), pp. 630-637.

[25] M. H. Stone, "Linear transformations in Hilbert space," *Proceedings of the National Academy of Sciences*, vol. 16 (1930), pp. 137-139. Previously, the theory of the discrete, instead of continuous, unitary cyclic groups in Hilbert space, was developed by A. Wintner, "Zur Theorie der beschränkten Bilinearformen," *Mathematische Zeitschrift*, vol. 30 (1929), pp. 228-282. It was shown by S. Bochner ("Spektralzerlegung linearer Scharen unitärer Operatoren," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1933, p. 371), that the method applied there in the discrete case can easily be transcribed to the continuous case, and this approach appears to be the simplest among the known proofs of Stone's result.

$$(35) \qquad \bar{f}.U^t f = \int_{-\infty}^{\infty} e^{itx} d_x(\bar{f}.\Phi(x)f).$$

Hence, $\Phi$ is uniquely determined by the flow $T^t$, $-\infty < t < \infty$.

Correspondingly, the point spectrum of the flow can be defined to be the point spectrum in the sense of Hilbert, that is, the sequence

$$(36) \qquad X: \qquad x_1, x_2, \cdots$$

of those points $x_n$ at which the monotone function (33) of $x$ is discontinuous for at least one $f$. In particular, the flow will be said to have a *pure* point spectrum if the spectral form, (33), of the flow reduces to

$$(37) \qquad \bar{f}.\Phi(x)f = \sum_{x_n < x} \bar{f}.(\Phi(x_n + 0) - \Phi(x_n - 0))f,$$

where $x$ and $f$ are arbitrary and $x_n$ ranges over the point spectrum, $X$.

A connection between these notions, which refer to the whole of the flow, and notions involving individual stream lines in the flow can easily be established, as follows:

THEOREM 5. *If (33) denotes the spectral form of a measure-preserving, $(P, t)$-measurable flow $T^t$ on a space $S$ of finite measure, then*

$$(38) \qquad \int_S \phi_P{}^f(x)\, d_P S = \bar{f}.\Phi(x)f, \qquad -\infty < x < \infty,$$

*for every function $f(P)$ of class $(L^2)$ on $S$, where $\phi_P{}^f$ denotes, for almost all $P$, the periodogram of $f(T^tP)$, $-\infty < t < \infty$.*

COROLLARY. *A real number, $x_0$, is a discontinuity point of the monotone function $\phi_P{}^f$ of $x$ for at least one fixed $f$ and for a $P$-set of positive measure, if and only if $x_0$ is in the point spectrum, $X$, of the flow; furthermore, the flow has a pure point spectrum if and only if $\phi_P{}^f$ is, for every fixed $f$ and for almost all $P$, a step function of $x$.*

In order to prove Theorem 5, let $f(P)$ be a function of class $(L^2)$ and let $s$, $A$, $B$ be real numbers. Then, according to Fubini's theorem,

$$\int_S \int_A^B f(T^{s+t}P)\bar{f}(T^tP)\, dt\, d_P S = \int_A^B \int_S f(T^{s+t}P)\bar{f}(T^tP)\, d_P S\, dt.$$

Hence, on replacing in the inner integral on the right the integration variable $P$ by $T^tP$, where $t$ is fixed, one sees from

$$\int_A^B \int_S f(T^sP)\bar{f}(P)\, d_PS\, dt = (B-A) \int_S f(T^sP)\bar{f}(P)\, d_PS$$

that

$$\frac{1}{B-A} \int_S \int_A^B f(T^{s+t}P)\bar{f}(P)\, dt\, d_PS = \int_S f(T^sP)\bar{f}(P)\, d_PS.$$

It follows, therefore, from (27) that

$$\int_S c_P{}^f(s)\, d_PS = \int_S f(T^sP)\bar{f}(P)\, d_PS.$$

On applying (29) on the left and (32) on the right, one obtains

$$\int_S \int_{-\infty}^{\infty} e^{isx} d_x \phi_P{}^f(x)\, d_PS = \bar{f}.U^sf,$$

since $U^sf = f(T^sP)$, by the definition of $U^s$. Hence, on writing $t$ for $s$, one sees, by applying Fubini's theorem on the left and (35) on the right, that

$$\int_{-\infty}^{\infty} e^{itx} d_x \left( \int_S \phi_P{}^f(x)\, d_PS \right) = \int_{-\infty}^{\infty} e^{itx} d_x (\bar{f}.\Phi(x)f),$$

where $-\infty < t < \infty$; whence (38) follows from the uniqueness theorem of the Fourier-Stieltjes transform.

**11.** The flows which are almost periodic in the sense indicated at the beginning of § 10 can now be characterized as follows:

THEOREM 6. *Let $T^t$ be a measure-preserving, $(P, t)$-measurable flow on a space $S$ of finite measure, and let $q \geqq 1$. Then $f(T^tP)$, $-\infty < t < \infty$, is, for every function $f(P)$ of class $(L^q)$ on $S$ and for almost all $P$, almost periodic $(B^q)$ if and only if the flow has a pure point spectrum.*

In other words, *the flow transforms every Lebesgue integrable function on $S$ into a Besicovitch almost periodic function of $t$ for almost all $P$ if and only if the flow is Bohr almost periodic.* It is understood that by the Bohr

almost periodicity of the flow is meant the Bohr almost periodicity of the scalar function (35) of $t$, where $f = f(P)$ is any function of class $(L^2)$ on $S$. Then the equivalence of the last italicized statement with Theorem 6 is clear from the remark made at the end of § 2.

In order to prove Theorem 6, associate with every $f(P)$ on $S$ the sequence of functions $f_1(P), f_2(P), \cdots$ on $S$ which are defined by

$$(39) \qquad f_n(P) = \begin{cases} f(P), & \text{if } |f(P)| \leq n; \\ 0, & \text{if } |f(P)| > n. \end{cases}$$

It is then clear, either from Birkhoff's proof of his ergodic theorem or from the proof of Wiener's dominated ergodic theorem, that

$$(39 \text{ bis}) \qquad M_t\{|f(T^tP) - f_n(T^tP)|^q\} \to 0 \quad \text{as} \quad n \to \infty$$

holds for almost all $P$, if $f(P)$ is any function of class $(L^q)$ on $S$. Hence, if $f_n(T^tP)$ is almost periodic $(B^q)$ for every $n$ and for a fixed $P$ which is not in a certain set of measure zero, then $f(T^tP)$ is almost periodic $(B^q)$ for the same $P$. But every $f_n(T^tP)$ is, by (39), a bounded function of $t$ and is, therefore, either almost periodic $(B^q)$ for every $q \geqq 1$ or not almost periodic $(B^q)$ for any $q \geqq 1$. Hence, Theorem 6 is equivalent to the following statement: For every measurable, bounded function $f(P)$ on $S$, the function $f(T^tP)$ of $t$ is almost periodic $(B^2)$ for almost all $P$ if and only if the flow has a pure point spectrum. Since the truth of this statement is clear from Theorem 4 and from the Corollary of Theorem 5, the proof of Theorem 6 is complete.

In order to attempt a dynamical understanding of the actual content of Theorem 6, use will be made of Koopman's interpretation [26] of the point spectrum in terms of the "first integrals" of the flow $T^t$; an interpretation which holds also when the point spectrum is not pure. In fact, whether the point spectrum, (36), does or does not satisfy the condition, (37), of purity, a real number, $x$, is in the point spectrum if and only if $e^{ixt}$ is a characteristic number of the unitary form (35) for every $t$, i. e., if and only if there exists on $S$ a function $F = F(P)$ which is not 0 (almost everywhere on $S$), is of class $(L^2)$, and has the property that

$$(40) \qquad e^{ixt}F(P) = F(T^tP), \quad \text{where} \quad -\infty < t < \infty,$$

is an identity (for almost all $P$). A corresponding statement holds concerning the multiplicities of characteristic numbers. But if $x = 0$, then (40) reduces to

$$F(P) = F(T^tP), \qquad -\infty < t < \infty,$$

which means that $F(P)$ is a "first integral" of the flow $T^t$ (provided that a subset of

[26] B. O. Koopman, *loc. cit.*[23], p. 318.

$S$ of measure zero is disregarded), unless $F(P)$ is constant on $S$ (almost everywhere). If, on the other hand, $x \neq 0$, then, on assuming that $F(P)$ is regular enough to make

$$F^*(P) = \arg F(P), \text{ i. e., } \exp iF^*(P) = F(P)/|F(P)|,$$

a meaningful definition of a function $F^*$ of the position $P$ on $S$ (almost everywhere), one sees that (40), can be written in the form

$$\text{(40 bis)} \qquad F^*(T^tP) \equiv xt + F^*(P) \pmod{2\pi}, \quad \text{where} \quad -\infty < t < \infty.$$

But (40 bis) means that also $F^*(P)$ corresponds to a "first integral," namely, to one belonging to an "ignorable coördinate" leading to an "angular variable." This is the interpretation of Koopman, referred to before.

Suppose now that the flow on $S$ has a *pure* point spectrum. Then, and only in this case, the characteristic numbers $x$ and the corresponding characteristic functions $F(P)$ determine all unitary invariants of the flow on the Hilbert space $(L^2)$ of $S$. Accordingly, the flows of pure point spectrum can be thought of as characterized by the fact that their complete system of unitary invariants is expressible in terms of "first integrals" alone. But then Theorem 6 means that the "first integrals" of exactly those flows have this completeness property which determine "conditionally periodic" paths as "general (or, rather, generic) solutions."

This seems to agree with certain statements of physicists, which have never been motivated in a mathematically reasonable direction; actually, most of the problems in question are not even meaningful without definite topological assumptions in the large. Correspondingly, the curious fact expressed by the following theorem is likely to have topological implications in the dynamical case, where $S$ is a manifold and $T^t$ is a continuous flow on $S$.

THEOREM 7. *If a measure-preserving, $(P, t)$-measurable flow $T^t$ on a space $S$ of finite measure has the property that $f(T^tP)$, $-\infty < t < \infty$, is, for every function $f(P)$ of class $(L^2)$ on $S$ and for almost all $P$, almost periodic $(B^2)$, then all the Fourier exponents of $f(T^tP)$ are contained, for almost all $P$, in the point spectrum, $X$, of the flow, and so in a single enumerable set which is independent of $P$ (and also of $f$).*

Theorem 6 and the hypothesis of Theorem 7 imply that the flow has a pure point spectrum. Hence, Theorem 7 follows from Theorem 2 and from the Corollary of Theorem 5. In fact, a real number $x$ is called a Fourier exponent of an almost periodic function $f(t)$, $-\infty < t < \infty$, if the $x$-th Fourier constant of $f(t)$ does not vanish.

**12.** The results obtained will now be applied to the case of metrical transitivity. In order to simplify some of the formulae belonging to this case, it will be assumed that meas $S = 1$. This is only a normalization, involving no loss of generality, since the case meas $S = \infty$ was always excluded, while the case meas $S = 0$ is always trivial. Since meas $S = 1$, the flow $T^t$ on $S$ is metrically transitive if and only if

$$M_t\{f(T^tP)\} = \int_S f(P)\, d_PS$$

holds for almost all $P$, whenever $f(P)$ is of class $(L) = (L^1)$ on $S$.

THEOREM 8. *If the flow $T^t$ on $S$, where $\operatorname{meas} S = 1$, is metrically transitive, there exists for every function $f(P)$ of class $(L)$ on $S$ a subset, $Z$, of $S$ which is of measure zero and such that, unless $P$ is in $S - Z$,*

(i) *the $x$-th Fourier constant, $a_P{}^f(x)$, of the function $f(T^tP)$, $-\infty < t < \infty$, (which need not be almost periodic $(B)$ in this case) exists for every $x$ and determines an "intensity," $|a_P{}^f(x)|$, which is independent of $P$ (so that only the "phase," $\arg a_P{}^f(x)$, of the $x$-th "amplitude," $a_P{}^f(x)$, can vary with $P$), where $Z$ is independent of $x$;*

(ii) *the point spectrum, $X$, of the flow (which may or may not be a pure point spectrum) contains every Fourier exponent ("frequency") of $f(T^tP)$, $-\infty < t < \infty$, i. e., every real number $x$ satisfying $a_P{}^f(x) \neq 0$;*

(iii) *in the particular case where the function $f$ on $S$ is of class $(L^2)$, the correlation and intensity functions of $f(T^tP)$, $-\infty < t < \infty$, are explicitly given by*

$$\text{(I)} \qquad \phi_P{}^f(x) = \bar{f}.\Phi(x)f, \qquad -\infty < x < \infty$$

*and*

$$\text{(II)} \qquad |a_P{}^f(x)|^2 = \bar{f}.(\Phi(x+0) - \Phi(x-0))f, \qquad -\infty < x < \infty,$$

*where $\bar{f}.\Phi(x)f$ denotes the spectral form of the flow.*

In order to prove this theorem, it will be sufficient to establish the first part of (iii), i. e. (I), for every $f$ of class $(L^2)$, but (i), (ii) and the second part of (iii), i. e. (II), only under the assumption that the given function on $S$, instead of being of class $(L)$ and of class $(L^2)$ respectively, is bounded and measurable. In fact, if $f$ is of class $(L)$ and $f_n$ denotes the function (39), then an obvious adaptation of (39 bis) shows that, when $x$ is fixed,

$$\lim_{n\to\infty} a_P{}^{f_n}(x) = a_P{}^f(x) \text{ for almost all } P;$$

so that the statements (i), (ii) follow for the given function $f$ of class $(L)$, if they are granted for every $f_n$. If in addition $f$ is of class $(L^2)$, then, by (39),

$$\lim_{n\to\infty} |\bar{f}.f - \bar{f}_n.f_n| = 0;$$

so that (II) follows for the given $f$, if it is granted for every $f_n$.

Next, let $f$ be of class $(L^2)$. Then, as shown in the proof of the italicized statement at the beginning of § 9 (in the paragraph preceding Theorem 4), the periodogram $\phi_P{}^f$ (exists and) is independent of $P$ for almost all $P$. Hence, (I) follows from (38), where $\operatorname{meas} S = 1$, for every $f$ of class $(L^2)$.

Accordingly, the proof of Theorem 8 will be complete if one verifies (i), (ii) and (II) for every bounded $f(P)$. But (II) implies the whole of (ii)

and it also implies the second part of (i), i. e., the statement that only the phase of $a_P{}^f(x)$ can vary with $P$. Finally, (I) reduces (II) to the relation

$$|a_P{}^f(x)|^2 = \phi_P{}^f(x+0) - \phi_P{}^f(x-0).$$

Hence, in order to complete the proof of Theorem (8), it is sufficient to prove that there exists for every bounded measurable function, $f$, a zero set, $Z$, such that $a_P{}^f(x)$ exists for $-\infty < x < \infty$ and satisfies the last relation, unless $P$ is in $S-Z$. But the existence of such a $Z$ follows from (31) and from Theorems 2 and 3 in the same way as in the proof of Theorem 8; in fact, (I) implies that the enumerability hypothesis of Theorem 8 is satisfied.

The simplest illustration of Theorem 8 is supplied by the "strongly mixing" flows, considered by Koopman and von Neumann.[27] In fact, these flows can be characterized as those metrically transitive flows for which the point spectrum, $X$, consists of a single point, $x = 0$; so that the criterion of Theorem 6 for almost periodicity $(B)$ is violated in its full extent. The other extreme case, that is, the case of a metrically transitive flow with a pure point spectrum, is represented by any Kronecker-Weyl flow on a torus.

## Appendix.

As mentioned at the beginning of § 5, the almost periodicity $(B^2)$ of a measurable function $f(t)$, $-\infty < t < \infty$, has thus far been meant in the sense that there exists for $f(t)$ a sequence of trigonometric polynomials, say

$$(41)\qquad r_1(t), r_2(t), \cdots, r_n(t), \cdots,$$

such that

$$(42)\qquad M_t\{|f(t) - r_n(t)|^2\} \text{ exists for } n = 1, 2, \cdots$$

and

$$(43)\qquad M_t\{|f(t) - r_n(t)|^2\} \to 0 \quad \text{as} \quad n \to \infty,$$

where $M_t\{\cdots\}$ is meant in the following sense:

$$(44)\quad M_t\{g(t)\} = \lim_{\substack{-A\to\infty \\ B\to\infty}} \frac{1}{B-A} \int_A^B g(t)\,dt, \text{ i. e.,} = \lim_{A\to\pm\infty} \frac{1}{A} \int_0^A g(t)\,dt.$$

Almost periodicity $(B^2)$ in this sense is a sharper property than almost periodicity $(B^2)$ according to the original definition of Besicovitch.[28] In fact, he requires the existence of a sequence (41) of trigonometric polynomials

[27] B. O. Koopman and J. v. Neumann, "Dynamical systems of continuous spectra," *Proceedings of the National Academy of Sciences*, vol. 18 (1932), pp. 255-263.

[28] Cf. A. S. Besicovitch, *Almost Periodic Functions*, Cambridge, 1932.

satisfying (42) and (43) when $M_t\{\cdots\}$ is defined not by (44) but as a principal limit, i. e., by

$$M_t\{g(t)\} = \lim_{A\to\infty} \frac{1}{2A} \int_{-A}^{A} g(t)dt. \tag{45}$$

The replacement of (44) by (45) is essential, for instance, in the theory of the Riemann zeta-function,[29] since those theorems on ordinary Dirichlet series which center about the mean-value theorem of Schnee do not, in general, hold if (44) is required instead of the more inclusive definition, (45), of $M_t\{\cdots\}$. Needless to say, the proofs of Theorems 1, 2, 3 hold, without any change, if (44) is replaced by (45) in (2), (3), (4) and (42)-(43). The resulting wordings of Theorems 1, 2, 3 are then, of course, independent of their original wordings. On the other hand, Theorem 5 then expresses a weaker statement than its original wording.

The situation with regard to Theorem 5 becomes quite different if (44) is replaced not by (45) but by

$$M_t\{g(t)\} = \lim_{B-A\to\infty} \frac{1}{B-A} \int_{A}^{B} g(t)dt, \tag{46}$$

where the limit process consists of an indefinite increase of the *length* of the integration domain $A \leqq t \leqq B$; a process which is compatible with either $A \to \infty$ or $B \to -\infty$. Let $f(t)$ be called almost periodic ($W^2$) if there exists a sequence (41) of trigonometric polynomials satisfying (42) and (43) in the sense of (46). It is easy to see that $f(t)$ is almost periodic ($W^2$) if and only if it is almost periodic in the sense of Weyl.[30] The proofs of Theorems 1, 2, 3 hold without change if (44) is replaced by (46) in (2), (3), (4) and (42)-(43) and, correspondingly, ($B^2$) in Theorem 3 by ($W^2$).

On the other hand, *Theorems* 4, 5, 6 *become false if* (44) *is replaced by* (46) *and, correspondingly,* ($B^q$) *by* ($W^q$). This is implied by the fact that a measure-preserving, $(P, t)$-measurable flow $T^t$ on $S$ may or may not be of the unrestricted type, that is to say such that $M_t\{f(T^tP)\}$ exists, for every function $f(P)$ of class $(L)$ on $S$ and for almost all $P$, if $M_t\{\cdots\}$ is meant in the sense (46), instead of the restricted sense (44) of Birkhoff's ergodic theorem.

---

[29] Cf. A. Wintner, "The almost periodic behavior of $1/\zeta(1+it)$," *Duke Mathematical Journal*, vol. 2 (1936), pp. 443-446; "Riemann's hypothesis and almost periodic behavior," *Universidad Mayor de San Marcos*, Lima, vol. 61 (1939), pp. 575-585.

[30] H. Weyl, "Integralgleichungen und fastperiodische Funktionen," *Mathematische Annalen*, vol. 97 (1926), pp. 338-356.

ADDENDUM.*

Since this paper was written, it has been possible to decide the alternative formulated at the end of § 4. The answer to the question turns out to be affirmative. In other words, the assertion of Theorem 2 holds without the assumption

$$|f(t)| < \text{const.} \tag{47}$$

also:

THEOREM 2 bis. *If $f(t)$ has a correlation function, the $x$-th Fourier constant of $f(t)$ exists and vanishes for all those values of $x$ which are continuity points of the periodogram of $f(t)$.*

In order to simplify the formulae occurring in the proof, let (1), where $-\infty < t < \infty$, be replaced by

$$M_t\{g(t)\} = \lim_{u\to\infty} \frac{1}{u}\int_0^u g(t)\,dt, \tag{48}$$

where $0 \leqq t < \infty$. It is clear from the remark which follows (1), that this modification is unessential.

For sake of a further typographical simplification, let it be assumed that

$$f(t) \equiv 0 \quad \text{for} \quad 0 \leqq t \leqq 1. \tag{49}$$

This assumption involves no loss of generality, since none of the $t$-averages $M$ occurring in Theorem 2 bis is influenced by a change of the values of $f(t)$ on a bounded $t$-interval.

In view of (49), the function

$$F^x(t) = \frac{1}{t}\int_0^t e^{-ixv} f(v)\,dv \tag{50}$$

is bounded on every bounded $t$-interval. Since the existence of the correlation function, (3), implies that

$$M_t\{|f(t)|^2\} < \infty, \tag{51}$$

it follows from the Schwarz inequality that

$$|F^x(t)| < \text{const.} \tag{52}$$

for $0 \leqq t < \infty$ and for every $x$.

The proof of Theorem 2 was based [10] on a Tauberian fact, according to which $M_t\{g(t)\}$ exists and vanishes whenever

* Received May 2, 1941.

$$|g(t)| < \text{const.} \tag{53}$$

and

$$\int_0^\infty g(ut)\,\frac{\sin^2 t}{t^2}\,dt \to 0 \quad \text{as} \quad u \to \infty. \tag{54}$$

It was shown *loc. cit.*[10] that if $f(t)$ has a correlation function, and if $x$ is any continuity point of the periodogram of $f(t)$, then

$$\lim_{\epsilon\to\infty} \epsilon \int_0^\infty e^{-ixt} f(t) \left(\frac{\sin \epsilon t}{\epsilon t}\right)^2 dt = 0, \tag{55}$$

whether the additional restriction $|f(t)| < \text{const.}$ is or is not satisfied. In fact, (55) becomes identical with the last formula line of § 5, *loc. cit.*[10], if one replaces the sequence $\{a_n\}$, considered there, by the present case of a function, $f(t)$. The restriction $|f(t)| < \text{const.}$ was there used, not in the proof of (55), but only in the passage from (55) to

$$M_t\{e^{-ixt} f(t)\} = 0; \tag{56}$$

this passage having been based on the Tauberian theorem quoted before.

In order to prove Theorem 2 bis, it will now be shown that (51) and (55) imply (56) without the restriction $|f(t)| < \text{const.}$ also (it being understood that the *existence*, and not the vanishing, of the mean-value (56) is the essential part of the statement).

First, on writing $t$ for $\epsilon t$ in (55),

$$\int_0^\infty e^{-ixt/\epsilon} f\left(\frac{t}{\epsilon}\right) \frac{\sin^2 t}{t^2}\,dt \to 0 \quad \text{as} \quad \epsilon \to 0.$$

Hence, on placing $v = 1/\epsilon$ and averaging with respect to $v$ between $v = 0$ and a large $v (= u)$,

$$\frac{1}{u}\int_0^u \left(\int_0^\infty e^{-ixtv} f(vt)\,\frac{\sin^2 t}{t^2}\,dt\right) dv \to 0 \quad \text{as} \quad u \to \infty,$$

and so, by Fubini's theorem,

$$\int_0^\infty \left(\frac{1}{u}\int_0^u e^{-ixtv} f(vt)\,dv\right) \frac{\sin^2 t}{t^2}\,dt \to 0 \quad \text{as} \quad u \to \infty$$

(in fact, (49) and (51) imply that the last iterated integral, when written as a double integral, exists absolutely for every fixed $u > 0$). Since

$$\frac{1}{u}\int_0^u e^{-ixtv}f(vt)\,dv \equiv \frac{1}{ut}\int_0^{ut} e^{-ixt}f(v)\,dv$$

is, by (50), identical with $F^x(ut)$, it follows that

$$\int_0^\infty F^x(ut)\,\frac{\sin^2 t}{t^2}\,dt \to 0 \quad \text{as} \quad u \to \infty.$$

The last relation and (52) show that both conditions, (54) and (53), of the Tauberian theorem are satisfied by $g = F^x$. Hence, $M_t\{g(t)\}$ exists and vanishes for $g = F^x$. This means, by (48) and (50), that

$$\frac{1}{u}\int_0^u \frac{dt}{t}\int_0^t e^{-ixv}f(v)\,dv \to 0 \quad \text{as} \quad u \to \infty.$$

Since (49) and Fubini's theorem imply that

$$\int_0^u\left(\int_0^t \frac{e^{-ixv}f(v)}{t}\,dv\right)dt = \int_0^u\left(\int_t^u \frac{e^{-ixt}f(t)}{v}\,dv\right)dt,$$

and since

$$\int_t^u \frac{e^{-ixt}f(t)}{v}\,dv = e^{-ixt}f(t)\log\frac{u}{t},$$

it follows that

(57) $$\frac{1}{u}\int_0^u e^{-ixt}f(t)\log\frac{u}{t}\,dt \to 0,$$ [cf. (49)],

where $u \to \infty$.

The assertion of Theorem 2 bis is (56), that is, by (48),

(58) $$\frac{1}{u}\int_0^u e^{-ixt}f(t)\,dt \to 0.$$

While it is an obvious Abelian fact that (58) implies (57), it is clear that (57) in itself cannot imply (58). However, it will now be shown that (57) implies (58) in virtue of (51). This will complete the proof of Theorem 2 bis.

On placing $g(t) = e^{-ixt}f(t)$, where $x$ is fixed, one sees from (49) that (57) and (58) become

$$\frac{1}{u}\int_1^u g(t)\log\frac{u}{t}\,dt \to 0 \tag{59}$$

and

$$\frac{1}{u}\int_1^u g(t)\,dt \to 0 \tag{60}$$

respectively; while (51) becomes

$$M_t\{|g(t)|^2\} < \infty. \tag{61}$$

Thus all that is to be proved is that (59) implies (60), if (61) is satisfied. Actually, (59) *implies* (60) *whenever*

$$\overline{\lim_{u\to\infty}}\frac{1}{u}\int_1^u |g(t)|^2 dt < \infty \tag{62}$$

(and even if only

$$\overline{\lim_{u\to\infty}}\frac{1}{u}\int_1^u |g(t)|^p\,dt < \infty \tag{62 bis}$$

holds for some $p > 1$).

In order to prove this, let $\vartheta$ be a fixed number in the interval $0 < \vartheta < 1$. Then $u \to \infty$ is equivalent to $\vartheta u \to \infty$; so that, on writing $\vartheta u$ for $u$ in (59), one sees that

$$\frac{\log\vartheta}{\vartheta u}\int_1^{\vartheta u} g(t)\,dt + \frac{1}{\vartheta u}\int_1^{\vartheta u} g(t)\log\frac{u}{t}\,dt \to 0$$

as $u \to \infty$. Since

$$\frac{1}{\vartheta u}\int_1^{\vartheta u} g(t)\log\frac{u}{t}\,dt + \frac{1}{\vartheta u}\int_{\vartheta u}^{u} g(t)\log\frac{u}{t}\,dt \equiv \frac{1}{\vartheta u}\int_1^{u} g(t)\log\frac{u}{t}\,dt \to 0,$$

by (59), it follows that

$$\frac{\log\vartheta}{\vartheta u}\int_1^{\vartheta u} g(t)\,dt - \frac{1}{\vartheta u}\int_{\vartheta u}^{u} g(t)\log\frac{u}{t}\,dt \to 0;$$

so that, since

$$\left|\int_{\vartheta u}^{u} g(t)\log\frac{u}{t}\,dt\right| \leqq \int_{\vartheta u}^{u} |g(t)|\log\frac{u}{\vartheta u}\,dt = |\log\vartheta| \int_{\vartheta u}^{u} |g(t)|\,dt,$$

division by $|\log\vartheta|$ gives

$$\overline{\lim_{u\to\infty}}\left|\frac{1}{\vartheta u}\int_{1}^{\vartheta u} g(t)dt\right| \leqq \frac{1}{\vartheta}\,\overline{\lim_{u\to\infty}}\,\frac{1}{u}\int_{\vartheta u}^{u} |g(t)|\,dt.$$

But the upper limit on the left is independent of $\vartheta$; so that, on estimating the integral on the right by the Schwarz inequality

$$\int_{\vartheta u}^{u} |g(t)|\,dt \leqq \left(\int_{\vartheta u}^{u} 1^2 dt\right)^{\frac{1}{2}} \left(\int_{\vartheta u}^{u} |g(t)|^2\right)^{\frac{1}{2}} = u(1-\vartheta)^{\frac{1}{2}}\left(\frac{1}{u}\int_{\vartheta u}^{u} |g(t)|^2\,dt\right)^{\frac{1}{2}},$$

one obtains the inequality

$$\overline{\lim_{u\to\infty}}\left|\frac{1}{u}\int_{1}^{u} g(t)dt\right| \leqq \frac{(1-\vartheta)^{\frac{1}{2}}}{\vartheta}\,\overline{\lim_{u\to\infty}}\left(\frac{1}{u}\int_{1}^{u} |g(t)|^2\,dt\right)^{\frac{1}{2}}.$$

On letting here $\vartheta \to 1-0$, one sees from the assumption (62), that the proof of (60) is complete.

It is clear from this proof that (62) can be generalized to (62 bis), since the inequality of Schwarz can be replaced by that of Hölder.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
THE JOHNS HOPKINS UNIVERSITY.

# THE CONTINUITY OF THE AREA OF HARMONIC SURFACES AS A FUNCTION OF THE BOUNDARY REPRESENTATIONS.*

By Marston Morse and C. Tompkins.

**1. Introduction.** We are concerned with surfaces $S$ in a euclidean $m$-space of coördinates $(x_1, \cdots, x_m) = (x)$. We restrict ourselves to harmonic surfaces $x_i = x_i(u, v)$, that is to surfaces for which the functions $x_i(u, v)$ are harmonic functions of $(u, v)$. These functions are defined on a connected region $B$ in the $(u, v)$-plane, or on a Riemann surface spread over the $(u, v)$-plane. The region $B$ is bounded by a finite number of non-intersecting circles.

We begin with the case where $B$ is single-sheeted and is bounded by a single circle $C$. For simplicity we take $B$ as the unit disc $u^2 + v^2 < 1$. Let $S$ be a surface with $x_i(u, v)$ harmonic on $B$ and continuous on the closure of $B$. Let $(r, \theta)$ be ordinary polar coördinates in the $(u, v)$-plane. Set

$$x_i(\cos\theta, \sin\theta) = p_i(\theta) \qquad (i = 1, \cdots, m). \tag{1.1}$$

We assume that the functions $p_i(\theta)$ have bounded variations. We shall designate the set $p_i(\theta)$ $(i = 1, \cdots, m)$ by $p$.

Subject to the above restrictions we shall regard $p$ as a variable, more particularly as a point in a metric space $M$. To define the distance $pq$ between two points $p$ and $q$ in $M$, let $V_i(p)$ denote the total variation of $p_i(\theta)$ on the range $0 \leqq \theta \leqq 2\pi$. Relative to $q$, $V_i(q)$ is similarly defined. We set (cf. footnote 5 in [3])

$$pq = \sum_i \{| V_i(p) - V_i(q) | + \max | p_i(\theta) - q_i(\theta) |\}, \tag{1.2}$$

where the maximum is taken over the range $0 \leqq \theta \leqq 2\pi$. It is clear that $pq$ satisfies the usual axioms for a metric space. Let $\Omega(p)$ be the area of the harmonic surface defined by $p$. Our principal theorem is as follows.

*The area $\Omega(p)$ varies continuously with $p$ on $M$.*

In § 2 we shall give an example showing that this theorem is false if the distance $pq$ is defined by the formula

$$pq = \sum_i \max | p_i(\theta) - q_i(\theta) |. \tag{1.3}$$

In further appraising the meaning of this theorem, recall that the Dirichlet-Douglas integral (cf. [2])

* Received March 15, 1941.

$$D(p) = \tfrac{1}{2} \iint_B \sum_i \left\{ \left(\frac{\partial x_i}{\partial u}\right)^2 + \left(\frac{\partial x_i}{\partial v}\right)^2 \right\} dudv$$

is lower semi-continuous. However $D(p)$ is not in general continuous for $p$ on $M$, or even on subsets of $M$ for which $D(p)$ is finite. An example showing this will be given in § 2.

The functions $D(p)$ and $\Omega(p)$ are important instruments in the study of minimal surfaces. They are connected by the relation (see [1], p. 312)

$$\Omega(p) \leqq D(p),$$

with the equality holding if and only if $S$ is a minimal surface. When $S$ is minimal, $p$ is termed a *critical point* of $D(p)$. For critical points $p$ then

$$\Omega(p) = D(p).$$

We accordingly have the following corollary of our theorem.

*The integral $D(p)$ is continuous on the subset of critical points $p$ of $D(p)$.*

This corollary is more striking if one recalls the corresponding theorem for extremals in ordinary, regular, positive definite problems. Replace $p$, for example, by an arbitrary curve joining two fixed points on a closed regular surface with coördinates which are 4-fold differentiable with respect to the surface parameters $(u, v)$. Replace $D(p)$ by the length $L(p)$ of $p$. The distance between two curves $p$ and $q$ may be taken as that defined by Fréchet. The analogue of our corollary is the theorem that $L(p)$ is continuous on the subset of geodesic segments which join two fixed points of our surface.

Our theorem has been used in the paper [3] to lighten the conditions (cf. [4], [7]) under which minimal surfaces of unstable type are proved to exist. We shall make further applications of this theorem in the theory of minimal surfaces in the large.

The theorem is extended to surfaces of general topological types where $B$ is bounded by $n$ circular contours (cf. [6]) and is a region on a Riemann surface. In this general case the area is shown to depend continuously not only on the boundary vectors, but also on the radii and centers of the circles bounding $B$.

**2. Preliminary remarks and examples.** Let $S$ be a harmonic surface $x_i = x_i(u, v)$ defined over the region $B$, $r < 1$, with the boundary representation $x_i = p_i(\theta)$ as in (1.1). We refer to the metric space $M$ of boundary representations $p$ with distances $pq$ defined as in (1.2).

*Example* 1. There are many examples which show that $D(p)$ is not in

general continuous on $M$. We shall give an illuminating example of this type, illustrating properties of $D(p)$ not ordinarily shown in such examples.

To that end we shall define a sequence $p^n$ $(n = 1, 2, \cdots)$ of representations $p$ in the space $(x) = (x_1, x_2)$. We shall make use of parameter transformations $\alpha = \phi^n(\theta)$ in which $\phi^n(\theta)$ is continuous and increasing, with $\phi^n(\theta + 2\pi) \equiv \phi^n(\theta) + 2\pi$. In the $(\alpha, \theta)$-plane the graph of $\alpha = \phi^n(\theta)$ shall consist of three successive line segments with $\phi^n(0) = 0$. The middle line segment shall have its midpoint at $(\pi, \pi)$, a length of $1/n$, and a positive slope $k_n$ to be defined. The curve $p^n$ shall have the form

$$x_1 = \cos[\phi^n(\theta)], \qquad x_2 = \sin[\phi^n(\theta)] \qquad (0 \leqq \theta \leqq 2\pi).$$

If $k_n$ becomes infinite for a fixed $n$, $D(p^n)$ does likewise as is well-known (cf. [1], p. 300). We shall take $k_n$ so large that $D(p^n)$ exceeds $n$, so that $D(p^n)$ becomes infinite with $n$. Nevertheless $\phi^n(\theta)$ converges uniformly to $\theta$, while $p^n$ converges to the representation $q$,

$$x_1 = \cos\theta, \qquad x_2 = \sin\theta. \tag{2.1}$$

Observe that $D(q)$ is finite. Thus $p^n$ converges on $M$ to $q$, while $D(p^n)$ does not converge to $D(q)$.

*Example* 2. Let $M^*$ be the metric space of rectifiable curves $p$ with distances defined as in (1.3). We shall give an example showing that $\Omega(p)$ is not continuous on $M^*$. We shall make use of a result of Radó, [5], p. 92. We shall consider curves $p$ in the space of coördinates $(x, y, z)$. Let $C$ be the projection of $p$ on the $(x, y)$-plane. Every point $(x, y)$ not on $C$ has a definite integer $n(x, y)$ as its topological index with respect to $C$. With Radó define $N(x, y)$ as $|n(x, y)|$ if $(x, y)$ is not on $C$, and as 0 otherwise. Then according to Radó,

$$\Omega(p) \geqq \int\int N(x, y)\,dxdy, \tag{2.2}$$

the integral being taken over the whole plane in the sense of Lebesgue.

We shall give an example in which a circle $q$ of the form

$$x = \cos\theta, \qquad y = \sin\theta, \qquad z = 0 \tag{2.3}$$

is a limit of a sequence $p^n$ of curves in the space $(x, y, z)$, convergence being defined by the metric of $M^*$. To that end we choose a point $P$ on $q$, and replace a segment $t_n$ of $q$ with midpoint at $P$ by a curve with the end points of $t_n$ and with a spiral-shaped projection on the plane $z = 0$ so chosen that for the curve $p^n$ replacing $q$, $\Omega(p^n)$ exceeds $n$. This can be done in accordance with (2.2) in such fashion that the length of $t_n$ is finite, that the diameter

of $t_n$ tends to 0 with $n^{-1}$, and that $\Omega(p^n)$ is finite. It is clear that $p^n$ will converge to the circle $q$ in the sense of $M^*$, provided the parameter $\theta$ is properly assigned to the segments $t_n$. Thus $p^n$ converges to $q$ in the sense of $M^*$, while $\Omega(p^n)$ does not converge to $\Omega(q)$.

The lengths of the above curves $p^n$ are not bounded. This leads to the following question, the answer to which is not known to the authors. For curves $p$ of uniformly bounded length with a metric given by (1.3) is $\Omega(p)$ continuous?

We shall need a few facts concerning the spaces $M$ and $M^*$. Recall that the metrics of $M$ and $M^*$ are defined respectively by (1.2) and (1.3). Let $v_i(\alpha, p)$ denote the total variation of $p_i(\theta)$ on the interval $0 \leqq \theta \leqq \alpha$. For $p$ fixed, $v_i(\alpha, p)$ is continuous in $\alpha$. For $\alpha$ and $p$ both variable, $v_i(\alpha, p)$ is lower semi-continuous, convergence of $\alpha$ being ordinary, and convergence of $p$ being convergence on $M^*$. If $p$ converges to $q$ on $M$, $v_i(2\pi, p)$ converges to $v_i(2\pi, q)$ in accordance with the definition of $pq$ on $M$. We shall prove the following lemma.

LEMMA 2.1. *If $p$ converges to $q$ on $M$, $v_i(\theta, p)$ converges to $v_i(\theta, q)$ uniformly with respect to $\theta$ $(0 \leqq \theta \leqq 2\pi)$.*

Let $\bar{p}$ denote the representation $p_i(-\theta)$ $(i = 1, \cdots, m)$. If $\alpha$ and $\beta$ are non-negative and sum to $2\pi$,

$$v_i(\alpha, p) + v_i(\beta, \bar{p}) = v_i(2\pi, p),$$

so that

$$v_i(\alpha, p) = v_i(2\pi, p) - v_i(\beta, \bar{p}). \tag{2.4}$$

Now $v_i(\beta, \bar{p})$ is lower semi-continuous in its arguments so that $-v_i(\beta, \bar{p})$ is upper semi-continuous. Since $v_i(2\pi, p)$ is continuous in $p$, it follows from (2.4) that $v_i(\alpha, p)$ is upper semi-continuous in $(\alpha, p)$. But $v_i(\alpha, p)$ is also lower semi-continuous, and hence continuous.

The uniform convergence affirmed in the lemma follows from the usual cluster point argument taking the compactness of the interval $0 \leqq \theta \leqq 2\pi$ into account.

**3. The theorem for a single contour.** On $S$, $x_i = x_i(u, v)$. Set

$$x_i(r \cos \theta, r \sin \theta) = h_i(r, \theta), \qquad (i = 1, \cdots, m),$$

and recall that for $r < 1$,

$$h_i(r, \theta) = \frac{1}{2\pi} \int_0^{2\pi} p_i(\alpha) \frac{(1 - r^2)}{1 - 2r \cos(\theta - \alpha) + r^2} \, d\alpha. \tag{3.1}$$

Let $\Omega_\rho(p)$ denote the area of that part of $S$ on which $r \leqq \rho$. We suppose that $0 < \rho < 1$. Recall that

$$\Omega(p) = \lim_{\rho=1} \Omega_\rho(p). \tag{3.2}$$

When $\rho$ is fixed, $\Omega_\rho(p)$ is continuous in $p$ for $p$ on $M$, for the area is given by the classical formula

$$\Omega_\rho(p) = \int_0^{2\pi} \int_0^{\rho} [\sum_{i,j} \Delta^2_{ij}]^{\frac{1}{2}} dr d\theta, \qquad (i, j = 1, \cdots, m;\ i < j) \tag{3.3}$$

where

$$\Delta_{ij} = \frac{\partial h_i}{\partial r} \frac{\partial h_j}{\partial \theta} - \frac{\partial h_j}{\partial r} \frac{\partial h_i}{\partial \theta}, \tag{3.4}$$

and the partial derivatives involved vary continuously with $r$, $\theta$, and $p$ on $M$.

To establish the continuity of $\Omega(p)$ on $M$, we shall first obtain an estimate of

$$\Omega_\rho(p) - \Omega_\sigma(p) \qquad (0 < \sigma < \rho < 1) \tag{3.5}$$

for $\sigma$ and $\rho$ near 1. It follows from (3.3) that

$$\Omega_\rho(p) - \Omega_\sigma(p) \leqq \int_0^{2\pi} \int_\sigma^{\rho} \sum_{i,j} | \Delta_{ij} | \, dr d\theta \qquad (i, j = 1, \cdots, m;\ i < j). \tag{3.6}$$

It will be sufficient for our purposes to obtain an estimate of the integral

$$K_{ij}(\rho, \sigma, p) = \int_0^{2\pi} \int_\sigma^{\rho} | \Delta_{ij} | \, dr d\theta \qquad (0 < \sigma < \rho < 1). \tag{3.7}$$

To that end we use (3.1), and set

$$1 - 2r \cos(\theta - \alpha) + r^2 = g(\alpha, \theta, r). \tag{3.8}$$

Observe that for $r < 1$,

$$\frac{\partial h_i}{\partial \theta} = \frac{1}{2\pi} \int_0^{2\pi} p_i(\alpha) \frac{\partial}{\partial \theta}\left(\frac{1 - r^2}{g}\right) d\alpha = \frac{-1}{2\pi} \int_0^{2\pi} p_i(\alpha) \frac{\partial}{\partial \alpha}\left(\frac{1 - r^2}{g}\right) d\alpha.$$

According to the theory of Riemann-Stieltjes integrals (cf. [8], p. 539), we may integrate by parts and thus obtain

$$\frac{\partial h_i}{\partial \theta} = \frac{1}{2\pi} \int_0^{2\pi} \frac{(1 - r^2)}{g} \, dp_i(\alpha). \tag{3.9}$$

Differentiating under the integral sign in (3.1), we find that

$$\frac{\partial h_i}{\partial r} = \frac{1}{\pi} \int_0^{2\pi} p_i(\alpha) \frac{[(1 + r^2) \cos(\theta - \alpha) - 2r]}{g^2(\alpha, \theta, r)} \, d\alpha.$$

On integrating by parts with the aid of formula 305 of Pierce's tables (third revised edition), we see that

$$\frac{\partial h_i}{\partial r} = \frac{1}{\pi}\int_0^{2\pi} \frac{\sin(\theta-\alpha)}{g(\alpha,\theta,r)}\,dp_i(\alpha), \tag{3.10}$$

again a Riemann-Stieltjes integral.

We shall make use of (3.9) and (3.10) to obtain a formula for $|\Delta_{ij}|$. Upon replacing $i$ by $j$ in (3.9) and (3.10) when necessary, and at the same time replacing the parameter of integration $\alpha$ by $\beta$, we find that

$$\begin{aligned} &\Delta_{ij}(r,\theta) \\ &= \frac{1}{2\pi^2}\int_0^{2\pi}\int_0^{2\pi}(1-r^2)\,\frac{[\sin(\theta-\alpha)-\sin(\theta-\beta)]}{g(\alpha,\theta,r)g(\beta,\theta,r)}\,dp_i(\alpha)\,dp_j(\beta). \end{aligned} \tag{3.11}$$

Let the total variation of $p_i(\alpha)$ from 0 to $\alpha$ be denoted by $v_i(\alpha)$. Recall also that

$$\begin{aligned} |\sin(\theta-\alpha)-\sin(\theta-\beta)| &= 2\left|\sin\frac{\alpha-\beta}{2}\cos\left(\theta-\frac{\alpha+\beta}{2}\right)\right| \\ &\leqq 2\left|\sin\frac{\alpha-\beta}{2}\right|. \end{aligned} \tag{3.12}$$

With the aid of (3.12), (3.11) yields the relation

$$|\Delta_{ij}(r,\theta)| \leqq \frac{1}{\pi^2}\int_0^{2\pi}\int_0^{2\pi}\frac{(1-r^2)\left|\sin\left(\frac{\alpha-\beta}{2}\right)\right|}{g(\alpha,\theta,r)g(\beta,\theta,r)}\,dv_i(\alpha)\,dv_j(\beta). \tag{3.13}$$

To reduce (3.13) further, we shall set

$$1-2r\left|\cos\frac{\alpha-\beta}{2}\right| + r^2 = R(\alpha,\beta,r),$$

and prove two lemmas.

LEMMA 3.1. *For* $0 \leqq r < 1$,

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{(1-r^2)}{g(\alpha,\theta,r)g(\beta,\theta,r)}\,d\theta \leqq \frac{2}{R(\alpha,\beta,r)}. \tag{3.14}$$

First recall that

$$\begin{aligned} |\cos(\theta-\alpha)+\cos(\theta-\beta)| &= 2\left|\cos\left(\theta-\frac{\alpha+\beta}{2}\right)\cos\frac{\alpha-\beta}{2}\right| \\ &\leqq 2\left|\cos\frac{\alpha-\beta}{2}\right|. \end{aligned}$$

It follows from the definitions of the terms involved that

$$g(\alpha,\theta,r)+g(\beta,\theta,r) \geqq 2R(\alpha,\beta,r). \tag{3.15}$$

Holding $(\alpha, \beta, r)$ fast, let $I$ and $J$ be respectively the sets of values of $\theta$ on the interval $(0, 2\pi)$ for which

$$g(\alpha, \theta, r) \geqq R(\alpha, \beta, r), \tag{3.16}$$

and

$$g(\alpha, \theta, r) < R(\alpha, \beta, r). \tag{3.17}$$

On $J$, $g(\beta, \theta, r) > R(\alpha, \beta, r)$ by virtue of (3.15).

On $I$ and $J$ respectively,

$$1/g(\alpha, \theta, r) \leqq 1/R(\alpha, \beta, r), \qquad 1/g(\beta, \theta, r) < 1/R(\alpha, \beta, r).$$

Hence the integral in (3.14) is at most

$$\frac{1}{2\pi R} \int_I \frac{(1-r^2)}{g(\beta, \theta, r)} \, d\theta + \frac{1}{2\pi R} \int_J \frac{(1-r^2)}{g(\alpha, \theta, r)} \, d\theta. \tag{3.18}$$

But

$$\int_I \frac{(1-r^2)}{g(\beta, \theta, r)} \, d\theta \leqq \int_0^{2\pi} \frac{(1-r^2)}{g(\beta, \theta, r)} \, d\theta = 2\pi,$$

and a similar relation follows for the integral over $J$. Hence the sum (3.18) is at most $2R^{-1}$, and the proof of the lemma is complete.

By virtue of formula 67 of Pierce's tables, we find that for $0 < \sigma < \rho < 1$ and for $\alpha \neq \beta$,

$$\int_\sigma \frac{\left|\sin\left(\frac{\alpha-\beta}{2}\right)\right|}{1 - 2r\left|\cos\left(\frac{\alpha-\beta}{2}\right)\right| + r^2} \, dr = \left\{ \arctan\left( \frac{r - \left|\cos\left(\frac{\alpha-\beta}{2}\right)\right|}{\left|\sin\left(\frac{\alpha-\beta}{2}\right)\right|} \right) \right\}_{r=\sigma}^{r=\rho}. \tag{3.19}$$

We shall use this formula, representing the integral (3.19) as a function $Z(\sigma, \rho, \alpha, \beta)$. Referring to the left member of (3.19), we see that $Z(\sigma, \rho, \alpha, \beta)$ is continuous in its arguments for $0 < \sigma < \rho < 1$. We use that branch of the arctangent which ranges from $-\pi/2$ to $\pi/2$.

The following lemma is fundamental.

LEMMA 3.2. *Corresponding to an arbitrary curve $p^0$ of $M$ and a prescribed positive constant $e$, there exists a neighborhood $N$ of $p^0$ on $M$ and a constant $r_0 < 1$ so near 1 that for $p$ on $N$ and $r_0 < \sigma < \rho < 1$,*

$$\int_0^{2\pi} \int_\sigma^\rho |\Delta_{ij}(r, \theta)| \, dr d\theta < e. \tag{3.20}$$

Let $\eta$ be an arbitrary positive constant, and let $\delta$ be so small a positive constant that whenever $|\alpha - \alpha'| \leqq \delta$,

$$|v_i(\alpha) - v_i(\alpha')| < \eta \qquad (i = 1, \cdots, m) \tag{3.21}$$

for the variations $v_i(\alpha)$ and $v_i(\alpha')$ of $p_i{}^0(\alpha)$. Let $N$ be so small a neighborhood of $p^0$ on $M$ that (3.21) holds likewise for variations $v_i(\alpha)$ and $v_i(\alpha')$ of $p_i(\alpha)$ when $p$ is on $N$. That such a neighborhood exists follows from Lemma 2.1.

We substitute the right member of (3.13) for $|\Delta_{ij}|$ in (3.20), thereby replacing the integral (3.20) by an integral with respect to $\alpha$, $\beta$, $r$, and $\theta$. We invert the order of integration, first integrating with respect to $\theta$, using Lemma 3.1. Denoting the integral (3.20) by $K_{ij}(\sigma, \rho, p)$, we see that

$$\pi K_{ij} \leqq 4 \int_0^{2\pi} \int_0^{2\pi} \int_\sigma^\rho \frac{\left|\sin\left(\dfrac{\alpha-\beta}{2}\right)\right|}{R(\alpha,\beta,r)} dr dv_i(\alpha) dv_j(\beta).$$

We next integrate with respect to $r$, using (3.19). We find thereby that

$$\pi K_{ij} \leqq 4 \int_0^{2\pi} \int_0^{2\pi} Z(\sigma,\rho,\alpha,\beta) dv_i(\alpha) dv_j(\beta). \tag{3.22}$$

The function $Z(\sigma,\rho,\alpha,\beta)$ is periodic in $\alpha$ and $\beta$. For each $j$ $(j = 1, \cdots, m)$ and constant $h$, the difference

$$v_i(\alpha + h) - v_i(\alpha)$$

is likewise periodic in $\alpha$. The integral on the right of (3.22) may accordingly be given the form

$$4 \int_0^{2\pi} \int_{\beta-\pi}^{\beta+\pi} Z(\sigma,\rho,\alpha,\beta) dv_i(\alpha) dv_j(\beta).$$

For each $\beta$, the interval

$$\beta - \pi \leqq \alpha \leqq \beta + \pi \tag{3.23}$$

will be divided into the subinterval

$$\beta - \delta \leqq \alpha \leqq \beta + \delta \qquad (0 < \delta < \pi), \tag{3.24}$$

and its complementary set $I(\beta)$. We write the integral (3.22) in the form

$$4 \int_0^{2\pi} \int_{\beta-\delta}^{\beta+\delta} Z dv_i(\alpha) dv_j(\beta) + 4 \int_0^{2\pi} \int_{I(\beta)} Z dv_i(\alpha) dv_j(\beta). \tag{3.25}$$

We shall estimate the terms in (3.25). Let $B$ be a bound of $v_j(2\pi)$ as $p$ varies on $N$. Such a bound exists since (3.21) holds for $|\alpha - \alpha'| \leqq \delta$, and $p$ on $N$. Note that $Z$ is at most $\pi$. Hence

$$\begin{aligned} 4 \int_0^{2\pi} \int_{\beta-\delta}^{\beta+\delta} Z dv_i(\alpha) dv_j(\beta) &\leqq 4\pi \int_0^{2\pi} [v_j(\beta+\delta) - v_j(\beta-\delta)] dv_j(\beta) \\ &\leqq 4\pi(2\eta) B. \end{aligned}$$

For $\alpha$ on $I(\beta)$, we can give a sharper estimate of $Z$. For $\alpha$ on $I(\beta)$, $|\alpha-\beta| \geqq \delta$. If $r_0$ differs sufficiently little from 1 with $0 < r_0 < 1$, and if $r_0 < \sigma < \rho < 1$, then $Z < \eta$ [$\alpha$ on $I(\beta)$]. Hence

$$4 \int_0^{2\pi} \int_{I(\beta)} Z dv_i(\alpha) dv_j(\beta) \leqq 4\eta B^2.$$

Thus

$$\pi K_{ij}(\sigma, \rho, p) \leqq \eta[8\pi B + 4B^2]. \tag{3.26}$$

But $\eta$ is arbitrarily small and (3.26) holds for $p$ on $N$ and for $r_0 < \sigma < \rho < 1$, provided $N$ is sufficiently small and $r_0$ is sufficiently near 1, the choice of $N$ and $r_0$ depending on $\eta$ and $p^0$.

The lemma follows from (3.26).

We have not established the uniform convergence of $\Omega_\rho(p)$ as $\rho$ tends to 1, but the preceding lemma gives us a result equally effective in proving the continuity of $\Omega(p)$.

THEOREM 3.1. *The area $\Omega(p)$ of the harmonic surface of disc type defined by the boundary curve $p$ varies continuously with $p$, as $p$ varies on the metric space $M$.*

Recall that $m$ is the number of coördinates $x_i$. It follows from (3.6) and the preceding lemma that for $p$ on $N$,

$$\Omega_\rho(p) - \Omega_\sigma(p) < m^2 e \qquad (r_0 < \sigma < \rho < 1),$$

provided $N$ is a sufficiently small neighborhood of $p^0$ on $N$ and $r_0$ is sufficiently near 1. Letting $\rho$ tend to 1, we infer that

$$\Omega(p) - \Omega_\sigma(p) \leqq m^2 e \qquad (r_0 < \sigma < 1;\ p \in N). \tag{3.27}$$

For fixed $\sigma$ conditioned as in (3.27), let $N_1 \subset N$ be so small a neighborhood of $p^0$ that

$$|\Omega_\sigma(q) - \Omega_\sigma(p^0)| < e \qquad (q \in N_1).$$

Then

$$\begin{aligned} |\Omega(q) - \Omega(p^0)| &\leqq |\Omega(q) - \Omega_\sigma(q)| \\ &+ |\Omega_\sigma(q) - \Omega_\sigma(p^0)| + |\Omega_\sigma(p^0) - \Omega(p^0)| \leqq m^2 e + e + m^2 e. \end{aligned} \tag{3.28}$$

Thus the left member of (3.28) is arbitrarily small if $q$ lies on a sufficiently small neighborhood $N_1$ of $p^0$.

The proof of the theorem is complete.

**4. The general case.** We shall refer to various metric spaces $M_1, M_2, \cdots, M_r$ with points $X_1, X_2, \cdots, X_r$ respectively, and shall consider the corresponding product space $\Pi$, in which a point is determined by the ensemble

$(X_1, X_2, \cdots, X_r)$. Let $(X'_1, X'_2, \cdots, X'_r)$ be a second point in $\Pi$. With these two points in $\Pi$ we shall *associate* a distance given by

$$X_1X'_1 + X_2X'_2 + \cdots + X_rX'_r,$$

where $X_jX'_j$ is the distance in the $j$-th factor space. If the distances in the factor spaces satisfy the usual metric axioms, it is easy to prove that the distance associated with the product space satisfies these same axioms.

We return to the harmonic surface $S$. In the general case, the region of definition $B$ of the functions $x_i = x_i(u, v)$ defining $S$ is a bounded connected region of a multiply-sheeted Riemann surface $\Sigma$ spread over the $(u, v)$-plane. This Riemann surface shall possess at most a finite number of sheets and bear at most a finite number of branch points. The boundary of $B$ shall consist of a finite number of circles $C$ on $\Sigma$ which do not intersect on the Riemann surface nor pass through branch points. Let $(r, \theta)$ be polar coördinates in the $(u, v)$-plane in a system "belonging to $C$," that is with pole at the center of $C$ and with axis parallel to the positive $u$-axis. On each boundary circle $C$ there shall be represented a boundary curve $x_i = p_i(\theta)$. The functions $x_i(u, v)$ defining $S$ are harmonic over $B$ and continuous on $\bar{B}$. If $a$ is the radius of $C$, and $(\bar{u}, \bar{v})$ gives the center of $C$, we assume that

$$x_i(a \cos \theta + \bar{u}, a \sin \theta + \bar{v}) = p_i(\theta), \qquad (i = 1, \cdots, m),$$

where that branch of $x_i(u, v)$ is to be used which belongs to the sheet of $\Sigma$ which bears $C$. For simplicity we shall suppose from this point on that $\Sigma$ is a single-sheeted surface bounded by the various circles $C$.

Each boundary curve $p$ will be regarded as a point in a metric space $M$ with a distance function defined as in (1.2). The ensemble $\boldsymbol{p}$ of the respective boundary curves will be regarded as a point in a product space with its associated metric obtained as above from the metrics of its factors. The $(u, v)$-coördinates $(\bar{u}, \bar{v})$ of the centers of the boundary circles $C$ together with their radii will be called the *circle parameters* of the boundary circles. The ensemble $\boldsymbol{b}$ of these circle parameters will be regarded as a point in a euclidean space of the corresponding number of dimensions. We shall have occasion to regard the pair $(\boldsymbol{p}, \boldsymbol{b})$ as a point in a product space with the corresponding associated metric. Similarly, the ensemble $(u, v, \boldsymbol{p}, \boldsymbol{b})$ will be regarded as a point in a product space with corresponding associated metric.

The region $B$ and surface $S$ will be more precisely designated by $B(\boldsymbol{b})$ and $S(\boldsymbol{p}, \boldsymbol{b})$ respectively. The coördinates $x_i$ of the point on $S(\boldsymbol{p}, \boldsymbol{b})$ will then be represented by functions $f_i(u, v, \boldsymbol{p}, \boldsymbol{b})$ $(i = 1, \cdots, m)$. Let $C(\boldsymbol{b})$ designate an arbitrary boundary circle of $B(\boldsymbol{b})$, and let $x_i = p_i(\theta)$ be the

boundary curve defined over $C(\boldsymbol{b})$. If $(r,\theta)$ are polar coördinates belonging to $C(\boldsymbol{b})$, if $(\bar{u},\bar{v})$ gives the center of $C(\boldsymbol{b})$, and if $a$ is the radius of $C(\boldsymbol{b})$,

$$f_i(a\cos\theta+\bar{u}, a\sin\theta+\bar{v}, \boldsymbol{p}, \boldsymbol{b}) \equiv p_i(\theta) \qquad (i=1,\cdots,m).$$

The functions $f_i(u,v,\boldsymbol{p},\boldsymbol{b})$ are continuous in their arguments. This fact is most readily established by first showing that the Green's function $G$ set up for $B(\boldsymbol{b})$ with pole at $(u_0,v_0)$ is continuous on the product space of points $(u,v,u_0,v_0,\boldsymbol{b})$ for $(u,v)$ on $B$ and $(u,v)\neq(u_0,v_0)$. From this it follows by an appropriate local use of the Poisson integral that a directional derivative of $G$ on $B$ is a continuous function of the direction cosines involved and of the above arguments of $G$, provided $(u,v)\neq(u_0,v_0)$. Green's formula for $f_i$ then shows that $f_i$ depends continuously on its arguments.

We shall show that the area $\Omega(\boldsymbol{p},\boldsymbol{b})$ of $S(\boldsymbol{p},\boldsymbol{b})$ is continuous on the space of points $(\boldsymbol{p},\boldsymbol{b})$. Let $C(\boldsymbol{b})$ be one of the boundary circles of $B(\boldsymbol{b})$. Let $(\boldsymbol{p}_0,\boldsymbol{b}_0)$ be a particular point $(\boldsymbol{p},\boldsymbol{b})$. Let evaluation when $b \doteq b_0$ be indicated by adding the subscript 0. Let $C'$ be a circle on $B_0$, concentric with $C_0$ and so near $C_0$ that $C_0$ and $C'$ bound a plane ring region $R_0$ on $B_0$. We shall not vary $C'$ with $\boldsymbol{b}$. We shall vary $\boldsymbol{b}$ continuously among points so near $\boldsymbol{b}_0$ that $C'$ is bounded away from the boundary of $B(\boldsymbol{b})$. Let $R(\boldsymbol{b})$ be the region on $B(\boldsymbol{b})$ bounded by $C(\boldsymbol{b})$ and $C'$.

Recall that $a$ is the radius of $C(\boldsymbol{b})$. This radius $a$ is one of the coördinates of $(\boldsymbol{b})$. Let $(r,\theta)$ be polar coördinates belonging to $C(\boldsymbol{b})$. Suppose for simplicity that $R(\boldsymbol{b})$ is within $C(\boldsymbol{b})$. Set

$$(4.1)\quad \frac{1}{2\pi}\int_0^{2\pi} p_i(\alpha)\,\frac{(a^2-r^2)}{a^2-2ar\cos(\theta-\alpha)+r^2}\,d\alpha = h_i(r,\theta,\boldsymbol{p},\boldsymbol{b}), \qquad (r<a;\ i=1,\cdots,m).$$

On $R(\boldsymbol{b})$ set

$$(4.2)\qquad f_i(r\cos\theta+\bar{u}, r\sin\theta+\bar{v}, \boldsymbol{p}, \boldsymbol{b}) = \phi_i(r,\theta,\boldsymbol{p},\boldsymbol{b}).$$

The difference

$$(4.3)\qquad \phi_i - h_i = \omega_i(r,\theta,\boldsymbol{p},\boldsymbol{b})$$

is harmonic on $R(\boldsymbol{b})$ and, as $r$ tends to $a$, tends to 0 uniformly with respect to $\theta$. Hence $\omega_i$ may be harmonically continued over the ring region consisting of $R(\boldsymbol{b})$ and its reflection in $C(\boldsymbol{b})$. This continuation will again be denoted by $\omega_i$.

We wish to show that the area of the surface $S(\boldsymbol{p},\boldsymbol{b})$ varies continuously with $(\boldsymbol{p},\boldsymbol{b})$. If $H$ is a subregion of $B(\boldsymbol{b})$, we shall refer to the area of that part of $S(\boldsymbol{p},\boldsymbol{b})$ on which $(u,v)$ ranges over $H$ as the area of $S(\boldsymbol{p},\boldsymbol{b})$ *over* $H$. To establish the continuity of $\Omega(\boldsymbol{p},\boldsymbol{b})$, it will be sufficient to prove the continuity of the area of $S(\boldsymbol{p},\boldsymbol{b})$ over regions such as $R(\boldsymbol{b})$ neighboring the

respective circular boundaries of $B(\boldsymbol{b})$. For the area of $S(\boldsymbol{p},\boldsymbol{b})$ over the residue of $B(\boldsymbol{b})$ depends continuously on $(\boldsymbol{p},\boldsymbol{b})$ in an obvious manner.

We find it convenient to set

$$\Delta_{ij}(U,V) = \begin{vmatrix} \dfrac{\partial U_i}{\partial r} & \dfrac{\partial V_j}{\partial r} \\ \dfrac{\partial U_i}{\partial \theta} & \dfrac{\partial V_j}{\partial \theta} \end{vmatrix},$$

as a matter of formal notation. To show that the area of $S(\boldsymbol{p},\boldsymbol{b})$ over $R(\boldsymbol{b})$ varies continuously with $(\boldsymbol{p},\boldsymbol{b})$, it will be sufficient, as in § 3, to prove the following lemma.

LEMMA 4. 1. *Corresponding to an arbitrary positive constant $e$,*

$$\int_0^{2\pi}\int_\sigma^\rho |\Delta_{ij}(\phi,\phi)|\,drd\theta < e \qquad (a-\eta<\sigma<\rho<a), \tag{4. 4}$$

*provided $(\boldsymbol{p},\boldsymbol{b})$ ranges on a sufficiently small neighborhood of $(\boldsymbol{p}_0,\boldsymbol{b}_0)$, and $\eta$ is a sufficiently small positive constant.*

In this lemma $a$ is a coördinate of $\boldsymbol{b}$ and varies with $\boldsymbol{b}$.

Recalling that $\phi_i \equiv h_i + \omega_i$, we see that

$$\Delta_{ij}(\phi,\phi) = \Delta_{ij}(h,h) + \Delta_{ij}(h,\omega) + \Delta_{ij}(\omega,h) + \Delta_{jj}(\omega,\omega). \tag{4. 5}$$

Our lemma would hold were $\phi$ replaced by $h$, by virtue of Lemma 3. 2. It also holds if $\phi$ is replaced by $\omega$, since the partial derivatives of $\omega$ as to $r$ and $\theta$ are bounded for $(u,v)$ on $R(\boldsymbol{b})$ and $(\boldsymbol{p},\boldsymbol{b})$ on a sufficiently small neighborhood of $(\boldsymbol{p}_0,\boldsymbol{b}_0)$. Hence it will be sufficient to show that

$$\int_0^{2\pi}\int_\sigma^\rho \{|\Delta_{ij}(h,\omega)| + |\Delta_{ij}(\omega,h)|\}drd\theta < e, \tag{4. 6}$$

subject to the conditions of the lemma. This in turn will follow provided we establish the following lemma.

LEMMA 4. 2. *Corresponding to an arbitrary positive constant $e_1$,*

$$\int_0^{2\pi}\int_\sigma^\rho \left|\frac{\partial h_i}{\partial\theta}\,\frac{\partial\omega_j}{\partial r}\right| drd\theta < e_1, \qquad (a-\eta<\sigma<\rho<a), \tag{4. 7}$$

and

$$\int_0^{2\pi}\int_\sigma^\rho \left|\frac{\partial h_i}{\partial r}\,\frac{\partial\omega_j}{\partial \theta}\right| drd\theta < e_1, \qquad (a-\eta<\sigma<\rho<a), \tag{4. 8}$$

*provided $(\boldsymbol{p},\boldsymbol{b})$ ranges over a sufficiently small neighborhood of $(\boldsymbol{p}_0,\boldsymbol{b}_0)$ and $\eta$ is a sufficiently small positive constant.*

For $(\boldsymbol{p}, \boldsymbol{b})$ on a sufficiently small neighborhood $N$ of $(\boldsymbol{p}_0, \boldsymbol{b}_0)$ and for $(u, v)$ on $R(\boldsymbol{b})$, the quantities

$$\left|\frac{\partial \omega_j}{\partial r}\right|, \qquad \left|\frac{\partial \omega_j}{\partial \theta}\right|,$$

admit a bound $K$. As in (3.9) we find

$$\frac{\partial h_i}{\partial \theta} = \frac{1}{2\pi}\int_0^{2\pi} \frac{(a^2 - r^2)}{a^2 - 2ar\cos(\theta - \alpha) + r^2}\, dp_i(\alpha).$$

Hence the integral in (4.7) is at most

$$\frac{K}{2\pi}\int_0^{2\pi}\int_\sigma^\rho\int_0^{2\pi} \frac{(a^2 - r^2)}{a^2 - 2ar\cos(\theta - \alpha) + r^2}\, d\theta dr dv_i(\alpha) \leqq K\int_0^{2\pi}\int_\sigma^\rho dr dv_i(\alpha).$$

Hence (4.7) holds subject to the conditions of the lemma.

Let the integral in (4.8) be denoted by $J$. As in the proof of (3.10), we find that

$$\frac{\partial h_i}{\partial r} = \frac{a}{\pi}\int_0^{2\pi} \frac{\sin(\theta - \alpha)}{a^2 - 2ar\cos(\theta - \alpha) + r^2}\, dp_i(\alpha),$$

so that for $(\boldsymbol{p}, \boldsymbol{b})$ on $N$ and $(u, v)$ on $R(\boldsymbol{b})$,

$$J \leqq \frac{Ka}{\pi}\int_0^{2\pi}\int_0^{2\pi}\int_\sigma^\rho \frac{|\sin(\theta - \alpha)|}{a^2 - 2ar\,|\cos(\theta - \alpha)| + r^2}\, dr d\theta dv_i(\alpha).$$

Referring to formula 67 of Pierce's tables, we see that

$$J \leqq \frac{Ka}{\pi}\int_0^{2\pi}\int_0^{2\pi} W d\theta dv_i(\alpha), \tag{4.9}$$

where

$$W = \left[\frac{1}{a}\arctan\left\{\frac{r - a\,|\cos(\theta - \alpha)|}{a\,|\sin(\theta - \alpha)|}\right\}\right]_\sigma^\rho,$$

and where the arctangent may be taken between $-\pi/2$ and $\pi/2$ inclusive. Since $W$ has the period $\pi$ in $\theta$, (4.9) may be written in the form

$$J \leqq \frac{2Ka}{\pi}\int_0^{2\pi}\int_{\alpha-\pi/2}^{\alpha+\pi/2} W d\theta dv_i(\alpha). \tag{4.10}$$

Recall that $W \leqq \pi$, and let $\delta$ be an arbitrarily small positive constant $< \pi/2$. From (4.10) we infer that

$$J \leqq 2Ka\int_0^{2\pi}\int_{\alpha-\delta}^{\alpha+\delta} d\theta dv_i(\alpha) + \frac{2Ka}{\pi}\int_0^{2\pi}\int_{I(\delta)} W d\theta dv_i(\alpha), \tag{4.11}$$

where $I(\delta)$ is complementary to the interval $(\alpha - \delta, \alpha + \delta)$ on the interval $(\alpha - \pi/2, \alpha + \pi/2)$. Let $\boldsymbol{p}$ be restricted to so small a neighborhood of $\boldsymbol{p}_0$

that $v_i(2\pi)$ admits a bound $L$, independent of $\boldsymbol{p}$. Let $\eta > 0$ be so small that when $\theta$ is on $I(\delta)$ and $a-\eta < \rho < \sigma < a$, $W$ is at most $\delta$. Then the sum in (4.11) is at most

$$4KaL\delta + 4Ka\delta L \geqq J.$$

Since $\delta$ is arbitrarily small, the integral $J$ is arbitrarily small, subject to the conditions of the lemma.

The proof of the lemma is complete.

Lemma 4.1 follows from Lemma 4.2 as we have already pointed out. From Lemma 4.1 the continuity of the area of $S(\boldsymbol{p}, \boldsymbol{b})$ over $R(\boldsymbol{b})$ may be inferred essentially as in § 3. The continuity of the area $\Omega(\boldsymbol{p}, \boldsymbol{b})$ then follows as indicated prior to Lemma 4.1. Hence we have proved the following theorem.

THEOREM 4.1. *The area $\Omega(\boldsymbol{p}, \boldsymbol{b})$ of the harmonic surface $S(\boldsymbol{p}, \boldsymbol{b})$ assuming the boundary curves of the set $\boldsymbol{p}$ on the boundary circles determined by the set $\boldsymbol{b}$ varies continuously with the set $(\boldsymbol{p}, \boldsymbol{b})$.*

THE INSTITUTE FOR ADVANCED STUDY
AND
PRINCETON UNIVERSITY.

## BIBLIOGRAPHY.

[1]. Douglas, "Solution of the problem of Plateau," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 263-321.

[2]. ———, "Minimal surfaces of higher topological structure," *Annals of Mathematics*, vol. 40 (1939), pp. 205-298.

[3]. Morse and Tompkins, "Minimal surfaces of non-minimum type by a new mode of approximation," *Annals of Mathematics*, vol. 42 (1941).

[4]. ——— ———, "The existence of minimal surfaces of general critical types," *Annals of Mathematics*, vol. 40 (1939), pp. 443-472.

[5]. Radó, "On the problem of Plateau," *Ergebnisse der Mathematik und ihrer Grenzgebiete*, vol. 2 (1933).

[6]. Courant, "The existence of minimal surfaces of given topological structure under prescribed boundary conditions," *Acta Mathematica*, vol. 72 (1940), pp. 51-97.

[7]. Shiffman, "The Plateau problem for non-relative minima," *Annals of Mathematics*, vol. 40 (1939), pp. 834-854.

[8]. Hobson, "*The theory of functions of a real variable*," vol. 1, Cambridge University Press (1926).

# THE NON-LINEAR BOUNDARY VALUE PROBLEM OF THE BUCKLED PLATE.*

By K. O. Friedrichs and J. J. Stoker.

**Introduction.** In this paper we develop methods for solving a non-linear boundary value problem which has its origin in a problem in elasticity. The methods yield (Part I) a complete numerical solution of the problem; we obtain also (Part II) the relevant uniqueness, existence, and convergence theorems. Although we treat a specific problem, the basic principles of our methods could be applied to a considerable variety of non-linear problems.

Our problem concerns the buckling of a thin elastic plate under forces acting in the plane of the plate. While it is true that the lowest "critical" load at which buckling *begins* can be determined by solving a linear eigenvalue problem, the treatment of the buckling for loads beyond the critical load requires of necessity the description of the situation by a boundary value problem for non-linear differential equations.

Non-linear differential equations for the case of the thin plate have been derived by v. Kármán,[1] cf. [13]. They contain the linear biharmonic differential operator (of order four) and quadratic terms in the second derivatives.

We confine our discussion here to the special case of the circular plate under compressive forces at the boundary with *radial symmetry* assumed. In this case the v. Kármán equations reduce to a pair of ordinary non-linear differential equations, each of the second order. The boundary value problem considered in this paper concerns the latter pair of equations. It is found that the problem depends essentially upon one parameter $N$—the ratio of the pressure $\bar{p}$ applied at the edge to the lowest critical pressure $p^0$ at which buckling just begins.

In Part I we explain a procedure for complete *numerical* solution of our problem. We have carried out the numerical solution and will report here on some of the results.[2] Our procedure yields solutions for the entire range

* Received August 31, 1940.

[1] For general theories of non-linear elasticity see Trefftz [21], Biot [1], and Murnaghan [16]. The theory of Biot, which assumes small strains, appears to include that of v. Kármán as a special case; this theory has been applied to the buckling of thick plates. Murnaghan does not even make the assumption of small strains, and is successful in applying his theory to two special problems involving very large strains.

[2] A complete numerical and graphical discussion of these results, particularly those

$0 \leq N \leq \infty$; to cover this range we employ three methods which are different but which interlock. Each of the three methods is suitable for a particular range of values of $N$: 1. perturbation method (sec. 2) for low values of $N$, 2. power series method (sec. 3) for intermediate $N$, 3. asymptotic solution (sec. 4) for $N \to \infty$.

The perturbation method consists in developing the essential quantities with respect to a parameter and solving the sequence of linear boundary value problems which arise from the v. Kármán equations. The first such linear problem is nothing but the above mentioned eigenvalue problem, the lowest eigenvalue of which yields the critical load where bifurcation begins. This method is used for a rather low range of the ratio $N$, say $1 \leq N \leq 2.5$; beyond this range the amount of numerical computation involved is excessive.

The power series method consists in developing the essential quantities with respect to powers of the distance $r$ from the center and satisfying the boundary conditions by solving a transcendental equation. Certain peculiar difficulties encountered in solving the latter equation can be overcome easily by using results furnished by the perturbation method. In this way one can obtain solutions for a higher range of values of $N$, say $N \leq 15$, but hardly for larger values of $N$ since the labor of calculating again becomes excessive.

The asymptotic solution (sec. 4) gives the limit situation for $N \to \infty$. It is also characterized by a non-linear boundary value problem. The formulation of this problem is based on the occurrence of a boundary layer effect which in our computations had already become apparent for $N < 15$. The boundary layer effect can be roughly described as follows: While with increasing $N$ all quantities tend to become constant in the interior of the plate, they change rapidly in a narrow strip at the boundary.[3] Once the asymptotic solution has been found it is possible by a perturbation method to develop the solution in the neighborhood of $N = \infty$. In sec. 4 we carry out the first step in such a development.

In Part I (sec. 1 to 4) of this paper our problem is treated from the point of view of applied mathematics. In Part II, which is quite independent[4] of Part I, we investigate our problem with regard to existence and uniqueness of the solutions and their continuous dependence on $N$ for $N = \infty$. The

---

of practical significance, will appear, cf. [6]. A short note discussing our results has appeared, cf. [5].

[3] Such an edge effect presents some analogy to Prandtl's boundary layer phenomena encountered in connection with the flow of viscous fluids around obstacles (cf. e. g. [19]). This analogy aided us materially in finding the proper mode of attack for the asymptotic solution.

[4] However, some of the assumptions made in Part I receive justification in Part II.

discussion is based on a minimum problem and its relation to the boundary value problem.[5]

In sec. 5 these problems are completely formulated for the case of finite $N$. In sec. 7 the same problems are formulated in terms of new variables in such a way as to make possible simultaneous treatment of the finite ($N < \infty$) and asymptotic ($N = \infty$) cases. We prove in sec. 8 that the minimum problem has at most one solution, apart from sign, and that such a solution never vanishes; in addition we show that a solution of the boundary value problem which never vanishes solves the minimum problem. To prove the latter statement we show that the solution of our non-linear minimum problem, which is of the fourth degree, is at the same time the solution of a certain quadratic minimum problem to which Jacobi's transformation of the second variation can be applied. The existence of the solutions is treated in sec. 9 by direct methods; in addition, we prove the continuous dependence of the solution on $N$ for $N = \infty$, that is, we show that the solutions for finite $N$ converge to the asymptotic solution as $N \to \infty$. This asymptotic treatment, however, refers only to the boundary layer effect. In sec. 10 we give a rigorous treatment of the limit state in the *interior* of the plate and its connection with the boundary layer.

While it is true that the minimum problem has only one solution, apart from sign, the boundary value problem will have more and more solutions as $N$ increases. We show, however, in sec. 6 that the boundary value problem has at most three solutions, one identically zero, the others differing only in sign, provided that $N$ is not too large. The method used combines well-known facts concerning linear eigenvalue problems with geometrical reasoning and could be applied to a more general class of non-linear problems.

Appendix I is devoted to an investigation of the onset of buckling from the point of view of E. Schmidt's bifurcation theory (cf. e.g. [14]); this leads to a justification of some assumptions made in working with the perturbation method.

## PART I.

**1. Formulation of the problem.** We first introduce the v. Kármán equations (cf. 13), reference to which has been made in the introduction. They are the following pair of non-linear differential equations for two functions $\phi$ and $w$ of the variable $x$ and $y$:

$$(1.01) \qquad \nabla^4\phi = w_{xx}w_{yy} - w^2{}_{xy}, \quad \nabla^4 = \nabla^2\nabla^2, \quad \nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2,$$

$$(1.02) \qquad (\gamma h)^2\nabla^4 w + (\phi_{yy}w_{xx} - 2\phi_{xy}w_{xy} + \phi_{xx}w_{yy}) = 0.$$

[5] For a different class of non-linear boundary value problems and their treatment by minimum problems, see Hammerstein [10].

The quantity $w$ is the deflection of the middle surface of the plate and $\phi$ is the Airy stress function, from which the stresses in the middle surface (the "membrane" stresses) are derived from the relations

$$\sigma_{xx} = \phi_{yy}, \qquad \sigma_{xy} = -\phi_{xy}, \qquad \sigma_{yy} = \phi_{xx}, \tag{1.03}$$

implying that the equations of equilibrium are identically satisfied. Subscripts $x$, $y$, and $r$ on all quantities but stresses $\sigma$ and strains $\epsilon$ denote differentiation with respect to these variables. (What we call stresses here are to be understood as stresses per modulus of elasticity $E$). The stresses in the middle surface are also the averages of the stresses over the thickness of the plate. As a consequence of the hypothesis that the squares of the slopes of the middle surface are of the same order of magnitude as the strains in the middle surface, the following non-linear relations hold between the strains $\epsilon$, stresses $\sigma$, and displacements $u$, $v$, $w$ of the middle surface

$$\begin{aligned} \epsilon_{xx} &= u_x + \tfrac{1}{2}w_x^2 = -(1+\nu)\sigma_{xx} + \nu(\sigma_{xx}+\sigma_{yy}), \\ \epsilon_{xy} &= \tfrac{1}{2}(u_y + v_x) + \tfrac{1}{2}w_x w_y = -(1+\nu)\sigma_{xy}, \\ \epsilon_{yy} &= v_y + \tfrac{1}{2}w_y^2 = -(1+\nu)\sigma_{yy} + \nu(\sigma_{xx}+\sigma_{yy}). \end{aligned} \tag{1.04}$$

The expressions for the strains here differ by the quadratic terms in $w_x, w_y$ from the expressions of the linear theory. The quantity $\nu$ is Poisson's ratio. Equation (1.01), the compatibility equation, is obtained from the second and third members of (1.04) as the result of eliminating $u$ and $v$. Equation (1.02) is the same as that obtained from the usual linear bending theory of plates; in (1.02) $h$ is the thickness of the plate and $\gamma^2 = 1/12(1-\nu^2)$.

We shall take as boundary conditions on $\phi$ and $w$ those corresponding to the following physical situation: Only one external force is applied—a uniform normal pressure at the edge and in the plane of the plate. The plate is simply supported, i. e., the deflection at the edge is zero and no bending moment is applied there. Our methods can, however, be extended with obvious modifications to other cases, e. g., the case of the clamped plate.

The essential restriction of this paper is that we consider only the circular plate with radial symmetry. As domain of the variables $x$ and $y$ we take therefore the circle $r^2 = x^2 + y^2 \leq R^2$ and assume that $\phi$ and $w$ *depend upon* $r$ *alone*; in this case (1.01) and (1.02) become ordinary differential equations. We introduce the new functions $p = r^{-1}\phi_r$, $q = -Rr^{-1}w_r$, and the linear differential operator

$$G = R^2 r^{-3} \frac{d}{dr} r^3 \frac{d}{dr}. \tag{1.05}$$

The quantity $p$ represents physically the membrane stress on an element

perpendicular to the radius. The differential equations (1. 01) and (1. 02) become

$$\frac{d}{dr}\, r^2[Gp - \tfrac{1}{2}q^2] = 0; \qquad \frac{d}{dr}\, r^2[\eta^2 Gq + pq] = 0,$$

where $\eta = \gamma h/R$; integration of these equations yields

(1. 06) $\quad Gp - \tfrac{1}{2}q^2 = c_1/r^2, \qquad$ (1. 07) $\quad \eta^2 Gq + qp = c_2/r^2.$

Unless the constants $c_1$ and $c_2$ are zero, the original functions $\phi$ and $w$ could not have continuous fourth derivatives, since continuous fourth derivatives of $\phi$ and $w$ at $r = 0$ imply, on account of radial symmetry, continuous second derivatives of $p$ and $q$ with respect to $r$ and hence continuity of the left members of (1. 06) and (1. 07) at $r = 0$. We therefore assume $c_1 = c_2 = 0$ and take as *the fundamental differential equations*[6] for our investigation

(1. 08) $\quad Gp = \tfrac{1}{2}q^2, \qquad$ (1. 09) $\quad \eta^2 Gq + pq = 0.$

As boundary conditions at the center we take

(1. 10) $\quad \dfrac{dp}{dr} = p_r = 0, \qquad$ (1. 11) $\quad \dfrac{dq}{dr} = q_r = 0 \quad \text{for} \quad r = 0$

which must be satisfied, again because of symmetry, if $\phi$ and $w$ are to possess continuous fourth derivatives. The physical situation which we consider leads to the following conditions at the boundary:

(1. 12) $\qquad p = \bar{p} \quad \text{for} \quad r = R,$

where $\bar{p}$ is a prescribed positive constant;

(1. 13) $\quad B_\nu q(R) \equiv Rq_r(R) + (1 + \nu)q(R) = 0 \qquad (r = R),$

which implies vanishing of the bending moment at the edge.

The equations (1. 08) to (1. 13) constitute the formulation of the boundary value problem to be discussed in what follows.

Once the functions $p$ and $q$ have been determined, all quantities of interest from the physical point of view are easily obtained. In particular, the most important stresses, the circumferential membrane stress $p_c$ and the radial bending stress $p_b$ at the lower surface of the plate are given by

(1. 15) $\quad p_c = rp_r + p = B_0 p, \quad$ and $\quad$ (1. 16) $\quad p_b = 6\gamma\eta B_\nu q,$

where the operator $B_\nu$ is defined in (1. 13).

Our problem would seem to depend upon several parameters; but only

[6] These equations are simpler in form than those used by others, for example Nádai [17, p. 288] and Way [22], who work with quantities other than $p$ and $q$.

two are really essential, namely, Poisson's ratio $\nu$ and what might be called the "thrust ratio"

$$\lambda^2 = \bar{p}/\eta^2. \tag{1.17}$$

This becomes evident upon introduction of the quantities

$$p^* = p/\eta^2, \qquad q^* = q/\eta, \qquad r_* = r/R \tag{1.18}$$

into equations (1.08) to (1.14) as well as (1.05), which then assume the form

(1.05)* $$G = r_*^{-3}\frac{d}{dr_*}\left(r_*^3\frac{d}{dr_*}\right),$$

(1.08)* and (1.09)* $$Gp^* = \tfrac{1}{2}q^{*2}, \qquad Gq^* + p^*q^* = 0,$$

(1.10)* and (1.11)* $$p^*_{r_*} = q^*_{r_*} = 0 \quad \text{for} \quad r_* = 0,$$

(1.12)* and (1.13)* $$\bar{p}^* = \lambda^2, \qquad B_\nu q^* = 0 \quad \text{for} \quad r_* = 1.$$

(Note that the parameters $\lambda$ and $\nu$ occur only in the boundary conditions). It follows that two plates with the same values of $\lambda$ and $\nu$ possess solutions differing only by constant multipliers.

The material constant $\nu$ may have any value between zero and one; usually it is about 0.3. In general, variations in $\nu$ affect the solutions of problems in elasticity very little. We therefore took a convenient fixed value (0.318) for $\nu$ in our numerical calculations. Hence only $\lambda$ remains as the essential parameter (cf. [8]).

**2. Perturbation method.** Our boundary value problem has the obvious solution $q \equiv 0$, $p \equiv$ const. This is also the *only* solution for sufficiently small values of the parameter $\lambda$—a fact which we prove in sec. 6 (cf. Theorem 6.1). At a certain value $\lambda_0$ of $\lambda$ there will be a bifurcation; for $\lambda > \lambda_0$ two solutions for which $q \not\equiv 0$ will exist; these solutions differ only in the sign of $q$. The onset of such a "buckling" is usually treated by linearizing the v. Kármán equations (i. e., by setting the right member of (1.01), or, for us, of (1.08) equal to zero); the result is an eigenvalue problem for which the lowest eigenvalue is $\lambda_0$. For information respecting what occurs for $\lambda > \lambda_0$, it is necessary to solve the non-linear equations. Our problem can be solved for $\lambda$ beyond $\lambda_0$ by the perturbation method. One finds that the above linearization is nothing but the result of the first step of this method.

The perturbation method consists in developing $\bar{p}$, $p$, and $q$ with respect to a parameter $\epsilon$. Since solutions $(p, q)$ must appear in pairs $(p, q)$, $(p, -q)$, it is natural to assume developments of the following type: [7]

---

[7] That there exist solutions given by expansions of this type can be shown by the methods of the bifurcation theory of E. Schmidt (cf. App. I).

$$\begin{aligned} \bar{p} &= \bar{p}^0 + \epsilon^2\bar{p}^{(2)} + \epsilon^4\bar{p}^{(4)} + \cdots \\ p(r) &= p^0(r) + \epsilon^2 p^{(2)}(r) + \epsilon^4 p^{(4)}(r) + \cdots \\ q(r) &= \epsilon q^{(1)}(r) + \epsilon^3 q^{(3)}(r) + \epsilon^5 q^{(5)}(r) + \cdots . \end{aligned}$$

From $\bar{p}(\epsilon)$ we then obtain $\epsilon$ as a function of $\bar{p}$.

We substitute the above expansions into the differential equations (1. 08) and (1. 09) and the corresponding boundary conditions. By collecting terms of the same order in $\epsilon$ we obtain a sequence of linear differential equations for $p^0, q^{(1)}, p^{(2)}, q^{(3)}, \cdots$ accompanied by boundary conditions. For $p^0$ we find

$$L_0: \qquad Gp^0 = 0, \qquad p^0(R) = \bar{p}^0, \qquad p_r{}^0(0) = 0;$$

the only solution of $L_0$ is $p^0 \equiv \text{const.}$, i. e., $p^0 = \bar{p}^0$. The value of $\bar{p}^0$ is, however, not yet determined. The next equations are

$$L_1: \qquad \eta^2 Gq^{(1)} + p^0 q^{(1)} = 0, \qquad B_\nu q^{(1)}(R) = 0, \qquad q_r{}^{(1)}(0) = 0.$$

This homogeneous differential equation with homogeneous boundary conditions will have as sole solution $q^{(1)} \equiv 0$ except when $p^0$ is an eigenvalue. The system serves as an eigenvalue problem for fixing the value of $p^0 = \bar{p}^0$; it is the same problem as would result from the customary linear treatment of stability, in accordance with the above remark. The lowest eigenvalue

$$p^0 = \lambda_0{}^2\eta^2 \tag{2. 1}$$

characterizes the thrust at which buckling begins and the solution bifurcates.

We are now in a position to introduce the parameter $N$ by

$$N = \bar{p}/p^0 \doteq \lambda^2/\lambda_0{}^2; \tag{2. 2}$$

$N$ will be used frequently as the essential parameter (cf. end of sec. 1) in place of $\lambda$.

The solution $q^{(1)}$ of the differential equation in $L_1$ satisfying $q_r{}^{(1)}(0) = 0$ is easily found in terms of the Bessel function of order one:

$$q^{(1)} = aRr^{-1}J_1(\lambda_0 R^{-1} r), \tag{2. 3}$$

where $a$ is a constant which we may choose arbitrarily. The lowest value of $\lambda_0$ for which the boundary condition $B_\nu q^{(1)}(R) = 0$ is satisfied, is

$$\lambda_0 = 2.06, \text{ when } \nu = .318.^8 \tag{2. 4}$$

In this manner the lowest critical value $\lambda_0{}^2 = 4.2436$ of the thrust ratio is determined.

---

[8] This value of $\nu$ is a convenient and reasonable one; we have used it throughout our numerical calculations (cf. end of sec. 1).

The equations for $p^{(2)}$ are

$$L_2: \qquad Gp^{(2)} = \tfrac{1}{2}(q^{(1)})^2, \qquad p^{(2)}(R) = \bar{p}^{(2)}, \qquad p_r^{(2)}(0) = 0.$$

Using (1.08) and (1.10) we obtain

$$\frac{dp^{(2)}}{dr} = \tfrac{1}{2}r^{-3}R^{-2}\int_0^r (q^{(1)})^2 r^3 dr.^{9}$$

By a second integration one obtains $p^{(2)}$, except that $\bar{p}^{(2)}$ has not yet been determined. It will, however, be fixed in the next step.

The problem for $q^{(3)}$ is:

$$L_3: \qquad Gq^{(3)} + p^0 q^{(3)} = -\,p^{(2)}q^{(1)}, \qquad B_\nu q^{(3)}(R) = 0, \qquad q_r(0) = 0.$$

The differential equation here is non-homogeneous, but the corresponding homogeneous problem, which is the same as $L_1$, has a solution not identically zero; thus $L_3$ presents the exceptional case in which the non-homogeneous problem need not have a solution and, if it has, the solution is not unique. In order that $L_3$ possess a solution, the right-hand side $-\,p^{(2)}q^{(1)}$ must be orthogonal to the solution $q^{(1)}$ of the homogeneous problem. That is $p^{(2)}$ must satisfy the orthogonality condition

$$\int_0^R p^{(2)}(q^{(1)})^2 r^3 dr = 0.$$

This relation serves to determine the boundary value $\bar{p}^{(2)}$.

Having satisfied the preceding condition we are sure that $L_3$ has a solution. Let $\hat{q}^{(3)}$ be one such solution; it follows that all functions of the form $q^{(3)} = \hat{q}^{(3)} + \alpha q^{(1)}$ are also solutions. The constant $\alpha$ is, in principle, undetermined.[10] We may choose $\alpha$ arbitrarily, but we wish to choose it in such a way as to obtain a good approximation to the complete solution.

---

[9] While this integral is expressible in terms of Bessel functions:

$$\frac{dp^{(2)}}{dr} = (4\lambda_0)^{-1}a^2r^{-1}J_1^2(\lambda_0R^{-1}r) - J_0(\lambda_0R^{-1}r)J_2(\lambda_0R^{-1}r),$$

$p^{(2)}$ itself, unfortunately, appears not to be expressible in terms of known functions.

[10] This does not mean that the final solution is not uniquely determined, but refers to the fact that the parameter $\epsilon$ can be chosen in different ways. Instead of $\epsilon$ one might just as well have taken for parameter any *function* of $\epsilon$ of the form

$$\hat{\epsilon} = \epsilon + \alpha\epsilon^3 + \cdots.$$

The development

$$q = \hat{\epsilon}q^{(1)} + \hat{\epsilon}^3\hat{q}^{(3)} + \cdots$$

is then equivalent to

$$q = \epsilon q^{(1)} + \epsilon^3(\hat{q}^{(3)} + \alpha q^{(1)}) + \cdots.$$

The Rayleigh-Ritz method is a natural procedure for this purpose. By inserting $q = (\epsilon + \alpha\epsilon^3)q^{(1)} + \epsilon^3\hat{q}^{(3)}$ into the strain energy functional[11] $V(q)$ for fixed $p$ we obtain a function of $\alpha$ and $\epsilon$. Since $q$ can be characterized (cf. sec. 5) as the function minimizing $V(q)$, a suitable relation between $\alpha$ and $\epsilon$ can be obtained by minimizing $V$ considered as a function of $\alpha$. The resulting condition for $\alpha$, however, is a rather complicated cubic equation with coefficients depending on $\epsilon$; instead of solving this equation, we develop it in powers of $\epsilon$ and retain the term of lowest order, which incidentally is linear in $\alpha$. We choose for $\alpha$ the value obtained from this linear equation.[12]

Having fixed the constant $\alpha$, and thus $q^{(3)}$, we proceed to determine $p^{(4)}$ and $p^{(5)}$ from the equations:

$$L_4:\quad Gp^{(4)} = q^{(1)}q^{(3)},\quad p^{(4)}(R) = \bar{p}^{(4)},\quad p_r^{(4)}(0) = 0.$$

$$L_5:\quad \eta^2 Gq^{(5)} + p^0 q^{(5)} = -p^{(2)}q^{(3)} - p^{(4)}q^{(1)},\quad B_V q^{(5)}(R) = 0,\quad q_r^{(5)}(0) = 0.$$

Here we face the same problem as that encountered above: the boundary value $\bar{p}^{(4)}$ must be determined from the condition that $L_5$ have a solution, while $q^{(5)}$ is determined only within a multiple of $q^{(1)}$, which multiple may be chosen arbitrarily, or fixed by a Rayleigh-Ritz procedure as above.

The general type of equations to be solved is:

$$L_n:\qquad Gp^{(n)} = f^{(n)}(r),\qquad (n \text{ even}),$$

$$L_{n+1}:\qquad \eta^2 Gq^{(n+1)} + p^0 q^{(n+1)} = f^{(n+1)}(r),$$

where the expressions $f^{(n)}$ and $f^{(n+1)}$ are quadratic in previously determined functions $q^{(i)}$ and $p^{(j)}$ with $i < n$ and $j \leq n$. Equations of the type $L_n$ can always be solved by two quadratures, which is obviously not the case with

---

[11] The functional $V$ is defined and discussed in sec. 5.

[12] The linear equation is

$$\Big[2\int_0^R p_r^{(11)}p_r^{(33)}r^3dr + 4\int_0^R (p_r^{(13)})^2 r^3 dr - R^{-2}p^{(2)}\int_0^R q^{(3)})^2 r^3 dr\Big]\int_0^R (q^{(1)})^2 r^3 dr$$
$$-\Big[6\int_0^R p_r^{(11)}p_r^{(13)}r^3dr - R^{-2}\bar{p}^{(2)}\int_0^R q^{(1)}q^{(3)}r^3dr\Big]\int_0^R q^{(1)}q^{(3)}r^3dr = 0,$$

where

$$p_r^{(mn)} = (R^{-2}/2)\int_0^r q^{(m)}(r_1)q^{(n)}(r_1)r_1^3dr_1;\qquad m, n = 1, 3$$

and

$$q^{(3)} = \hat{q}^{(3)} + \alpha q^{(1)}.$$

We used this equation in our calculations with results which seemed to justify the labor involved in such a procedure.

$L_{n+1}$. However, $L_{n+1}$ can be transformed by introducing the new dependent variable $u^{(n)} = q^{(n+1)}/q^{(1)}$ and it then takes the form

$$r^{-3}\frac{d}{dr}r^3(q^{(1)})^2\frac{d}{dr}u^{(n)} = q^{(1)}f^{(n+1)}.$$

The first integration yields

$$\frac{du^{(n)}}{dr} = r^{-3}\int_0^r q^{(1)}f^{(n+1)}r^3dr/(q^{(1)})^2.$$

A second integration determines $u^{(n)}$ and consequently $q^{(n+1)}$. The boundary condition $B_\nu q^{(n+1)}(R) = 0$ is transformed into $u_r^{(n)} = 0$ for $r = R$, i. e., $\int_0^R q^{(1)}f^{(n+1)}r^3dr = 0$, and this relation holds automatically since it expresses the condition that $L_{n+1}$ have a solution satisfying the boundary conditions. We carried the solutions so far as to calculate $q^{(5)}$ and $p^{(6)}$.

Since all quadratures, after the first, appear not to be expressible in terms of known functions, we found it necessary to operate with power series. Although the integration of power series presents no difficulty, the obvious necessity for multiplication and division of one power series by another makes the numerical calculations very laborious.[13]

The rapidity of convergence of perturbation series is a matter of general significance, in view of the wide use of such methods in problems similar to ours [cf. 9, 15, 18].[14] In our problem it is possible to check the accuracy of the perturbation method by comparison with the results of a different method (cf. sec. 3). Our calculations show that the perturbation series converge satisfactorily only for a rather small range of values of the ratio $N = \bar{p}/p^0 = \lambda^2/\lambda_0^2$. If errors up to about 4% are permitted, $q$ may be calculated for $N = \bar{p}/p^0 = 1.15, 1.8, 2.5$ by using terms up to and including those of first, third, and fifth order respectively; $p$ may be calculated for $N = 1.4, 2.2, 2.8$ by using terms up to and including those of second, fourth, and sixth order respectively. For the stresses $p_b$ and $p_c$ (cf. (1.16) and (1.15)) the convergence is not quite so good. Beyond these values of $N$ the approximations become inaccurate rather quickly.

---

[13] We retained eight terms of each power series and found this just accurate enough to compute $q^{(5)}$ and $p^{(6)}$, with which our calculations ended.

[14] Marguerre [15], for instance, mentions that his calculations for the rectangular plate, which implied two perturbations, may perhaps be valid up to $N = 20$. In our simpler case we find that values of $N$ of such magnitudes cannot be treated with three perturbations. In fact, such values of $N$ lead to solutions already in the asymptotic range (cf. sec. 4).

**3. Power series method.** The solutions of equations (1.08) and (1.09) can be obtained as power series in $r$. All coefficients of the series could be fixed in terms of the first coefficient in each series, the latter being determined by the boundary conditions. This would involve solving very complicated transcendental equations. It is possible, however, to transform the problem in such a way that only one transcendental equation of relatively simple type need be solved.[15]

We introduce the new independent variable

$$\alpha = \mathrm{A}R^{-1}r, \qquad 0 \leq \alpha \leq \mathrm{A};$$

and the new functions

$$\begin{aligned} \pi &= \mathrm{A}^{-2}p^* = \mathrm{A}^{-2}p/\eta^2 \\ \kappa &= \mathrm{A}^{-2}q^* = \mathrm{A}^{-2}q/\eta, \quad (\text{cf. } (1.18)), \end{aligned} \tag{3.1}$$

where A is a parameter at our disposal. The differential equations (1.08) and (1.09) assume the form

$$\begin{aligned} \alpha^{-3}\frac{d}{d\alpha}\alpha^3\frac{d}{d\alpha}\pi &= \tfrac{1}{2}\kappa^2, \\ \alpha^{-3}\frac{d}{d\alpha}\alpha^3\frac{d}{d\alpha}\kappa + \pi\kappa &= 0 \end{aligned} \tag{3.2}$$

with boundary conditions

$$\left.\frac{d\pi}{d\alpha}\right|^0 = 0, \qquad \left.\frac{d\kappa}{d\alpha}\right|^0 = 0, \tag{3.3}$$

$$B_\nu\kappa = 0 \quad \text{for} \quad \alpha = \mathrm{A}, \tag{3.4}$$

$$\pi(\mathrm{A}) = \bar{\pi}. \tag{3.5}$$

Instead of prescribing $\bar{\pi}$ and A we proceed as follows: $\pi(0)$ and $\kappa(0)$ are chosen, A is determined by solving (3.4), and $\bar{\pi}$ is calculated from (3.5).

The power series for $\pi$ and $\kappa$ contain only even powers of $\alpha$, a conclusion that can be drawn from (3.2) and (3.3); we write, therefore,

$$\pi = \sum_{k=0}^{\infty}\pi_k\alpha^{2k}, \qquad \kappa = \sum_{k=0}^{\infty}\kappa_k\alpha^{2k}.$$

Upon insertion of these power series into (3.2) we obtain the following recursion formulas for $\pi_k$ and $\kappa_k$:

$$2k(2k+2)\pi_k = \tfrac{1}{2}\sum_{m+n=k-1}\kappa_m\kappa_n, \qquad 2k(2k+2)\kappa_k = -\sum_{m+n=k-1}\pi_m\kappa_n.$$

---

[15] The power series method (with a different arrangement) was used by Hencky [11] and Way [22] for the problem of the bending of circular plates under lateral loads or edge moments.

After having chosen $\pi(0) = \pi_0$ and $\kappa(0) = \kappa_0$ the successive coefficients are calculated by these formulas. Equation (3.4) becomes

$$\sum_{k=0}^{\infty} (2k+1+\nu)\kappa_k \alpha^{2k} = 0, \tag{3.6}$$

and must be solved for its lowest root[16] $\alpha = \mathrm{A}$; $p^*$ and $q^*$ are then calculated from (3.1).

By this method we obtained solutions for a much higher range of values of $N$ with very much less labor in numerical computation than would be required by the perturbation method. However, it was necessary to calculate a rather large number of coefficients in order to obtain sufficient accuracy for the higher values of $N$, e. g., for $N = 14.7$, the most extreme case calculated, we found thirty terms in each series barely sufficient.

In applying the power series method for the higher values of $N$, the following considerations are essential. We wish to obtain a fairly even distribution for the values of $N$ and at the same time we want to know roughly where the solutions $\alpha = \mathrm{A}$ of equations (3.6) will be. The amount of labor in computation reduces to a minimum if $\mathrm{A} \sim 1$: for then we have a means of knowing the accuracy with which the coefficients need be calculated and also the number of terms to be computed in the series. In order to obtain $\mathrm{A} \sim 1$ and a pre-determined distribution of the values of $N$, it is necessary to make in advance fairly accurate estimates of the values of $p^* = p/\eta^2$ and $q^* = q/\eta$ at the center ($r = 0$). How to do this is not obvious, for the values of $p$ and $q$ at the center change with increasing $N$ in rather surprising and unforeseen ways, as our computations showed. This is one disadvantage of the power series method which, however, can be overcome by beginning with the perturbation method (which requires no estimates in advance), and pursuing the latter until the trend of the solutions with increasing $N$ becomes apparent. The solutions obtained above by the perturbation method for $N \leqq 2.5$ proved in fact to be amply sufficient for this purpose, although not all of the distinctive qualitative features of the solutions, in their dependence upon $N$, had yet appeared in this range.

In comparing the perturbation and power series methods, it should be mentioned that the former is applicable in principle to similar problems without radial symmetry; this appears not to be the case with the power series method.

**4. Asymptotic solution.** Although the power series method can be

[16] If we were to take the second root of (3.6) for A, we would find the prolongation into the non-linear range of the *second* eigenfunction of the linearized problem $L_1$ (sec. 2).

applied to solve our problem for rather high values of $N$, it will not serve to determine what occurs when $N$ tends to infinity, or, what is the same thing, when the thrust ratio $\lambda^2 = \bar{p}/\eta^2$ tends to infinity. Such a passage to the limit may be achieved physically in various ways which are mathematically equivalent: for example, one might take a fixed plate ($\eta$ fixed) and allow $\bar{p}$ to increase indefinitely, or hold $\bar{p}$ fixed and consider plates with slenderness ratios $\eta$ tending to zero.

In order to determine the asymptotic behavior of the solutions, it is necessary to formulate a limit boundary value problem by a passage to the limit in the original differential equations and boundary conditions. A simple and rather natural procedure would be to hold $\bar{p}$ fixed and let $\eta$ tend to 0 in equations (1.08) and (1.09) which then take the form

$$Gp = \tfrac{1}{2}q^2, \qquad pq = 0. \tag{4.01}$$

The only solution of these equations which satisfies the regularity conditions (1.10) and (1.11) is $q \equiv 0$, $p \equiv$ constant. One is then tempted to fix this constant by setting it equal to the prescribed value $\bar{p}$ at the edge. This means that in the limit there would be a hydrostatic compression ($p > 0$) throughout the plate. Such a procedure corresponds to the treatment of laterally loaded clamped sheets by Hencky [11, 12], cf. also [4, 2], and a similar method may well be legitimate in cases where no edge *compression* is prescribed. Applied to our case, however, wrong results would be obtained. The correct limit procedure can be found only by a deeper analysis of the nature of the solutions.

In our case we have found by numerical calculation (cf. [5] and [6]) that with increasing thrust ratio the membrane stress $p$ approaches a state of constant tension ($p < 0$) over an increasingly large part of the interior of the plate and that the transition from tension in the interior to the prescribed compression $\bar{p} > 0$ at the edge takes place in a narrow strip, the breadth of which decreases with increasing $\lambda$. (We shall use the parameter $\lambda$ instead of $N$ from now on).

These results of the numerical calculation indicate strongly the nature of the limit situation. In the interior of the plate, the above solution $q \equiv 0$, $p \equiv$ const. of (4.01) seems valid, but the constant should not be determined by setting it equal to $\bar{p}$. The constant can be fixed only by an investigation of the transition phenomena in the "boundary layer." The boundary layer phenomena are coupled with the fact that the order of the system of differential equations (1.08) and (1.09) has been reduced from four to two on passing to (4.01). It is not to be expected that the solutions of the limit system (4.01) can satisfy four boundary conditions. The above discussion indicates that the lost boundary conditions are those at the edge.

A treatment of such an edge effect requires that the scale be stretched with increasing $\lambda$ in such a manner that the width of the edge strip, or boundary layer (as measured in the new scale) does not shrink to zero. This will be accomplished, as we shall see, by introducing the new independent variable

$$\beta = \lambda(1 - r/R), \qquad 0 \leq \beta \leq \lambda, \tag{4.02}$$

where

$$\lambda = \bar{p}^{\frac{1}{2}}/\eta = \lambda_0 \sqrt{N}. \tag{4.03}$$

Subscripts $\beta$ in what follows denote differentiation with respect to $\beta$. Upon introduction of

$$P = p/\bar{p}, \qquad Q = \eta q/\bar{p} \tag{4.04}$$

as new dependent variables, the original differential equations (1.08) and (1.09) take the forms:

$$\begin{aligned} P_{\beta\beta} - [3/(\lambda - \beta)]P_\beta &= Q^2/2 \\ Q_{\beta\beta} - [3/(\lambda - \beta)]Q_\beta + PQ &= 0. \end{aligned} \tag{4.05}$$

The boundary conditions (1.12) and (1.13) become

$$P(0) = 1, \qquad Q_\beta(0) - [(1 + \nu)/\lambda]Q(0) = 0, \tag{4.06}$$

while the regularity conditions (1.10) and (1.11) take the form

$$P_\beta(\lambda) = Q_\beta(\lambda) = 0. \tag{4.07}$$

We now let $\lambda$ tend to infinity and obtain the limit differential equations

$$P_{\beta\beta} = Q^2/2 \tag{4.08}$$

$$Q_{\beta\beta} + PQ = 0; \tag{4.09}$$

$P_\beta$

the boundary conditions (4.06) become

(4.10) $Q_\beta(0) = 0$, (4.11) $P(0) = 1$, (4.12) $Q_\beta(\infty) = P(\infty) = 0$.

The equations (4.08) to (4.12) constitute the formulation of the limit boundary value problem concerned with the "boundary layer." (This is not to be confused with the limit problem partially formulated in (4.01), which is concerned with the interior of the plate.) The present problem has the trivial solution $P = 1$, $Q = 0$; if it has another solution $P$, $Q \not\equiv 0$, then also $P$, $-Q$ is a solution. Hence the sign of $Q$ is arbitrary.

The equations (4.08) and (4.09) possess the first integral:

$$Q_\beta^2 - P_\beta^2 + PQ^2 = \text{const.} \tag{4.13}$$

It is plausible to take zero as the value of the constant. This, in view of the boundary conditions (4.12), is equivalent to assuming $PQ^2 = 0$ at $\beta = \infty$. (For a justification of this assumption see [34]). We take, therefore,

$$Q_\beta^2 - P_\beta^2 + PQ^2 = 0. \tag{4.14}$$

In view of (4.10) and (4.11) we find $P_\beta^2 = Q^2$ for $\beta = 0$, or $P_\beta = \pm Q$. Since we may choose either sign for $Q$, we set

$$P_\beta(0) = -Q(0). \tag{4.15}$$

We introduce new variables $x$, $y$, $z$ (not to be confused with the space variables used earlier) as follows:

$$x = \xi e^{-\omega\beta}, \qquad 0 \leq x \leq \xi, \tag{4.16}$$

$$y = -\omega^{-2}P, \tag{4.17}$$

$$z = \tfrac{1}{2}2^{\frac{1}{2}}\omega^{-2}Q, \tag{4.18}$$

where $\xi$ and $\omega$ are numbers to be determined. The differential equations (4.08) and (4.09) in these variables are:

$$x(xy_x)_x + z^2 = 0, \tag{4.19}$$

$$x(xz_x)_x - yz = 0. \tag{4.20}$$

The introduction of the new variable $x$ has the effect that the resulting differential equations (4.19) and (4.20) possess solutions expressible as power series in $x$:

$$y = \sum_{k=0}^{\infty} (-1)^k y_k x^{2k} \tag{4.21}$$

$$z = \sum_{m=0}^{\infty} (-1)^m z_m x^{2m+1}. \tag{4.22}$$

Upon substituting (4.21) and (4.22) into (4.19) and (4.20) we find the following formulas for $y_k$ and $z_m$:

$$(2k)^2 y_k = \sum_{m+n=k-1} z_m z_n, \tag{4.23}$$

$$(2m+1)^2 z_m = \sum_{n+k=m} z_n y_k. \tag{4.24}$$

There is only one arbitrary coefficient, namely $z_0$, to which we assigned the numerical value $z_0 = 4$. The coefficient $y_0$ is determined from (4.24) for $m = 0$ and has obviously the value $y_0 = 1$. This fact makes it possible to re-write (4.24) as a proper recursion formula:

$$2m(2m+2)z_m = \sum_{n=0}^{m-1} z_n y_{m-n}. \tag{4.25}$$

We found it amply sufficient to compute ten terms in each series.

We turn now to consideration of the boundary condition associated with (4.19) and (4.20). The regularity conditions (4.12) are satisfied automatically in view of (4.16), (4.17), (4.18), and the assumed development into the power series (4.21) and (4.22). The boundary conditions (4.10) and (4.11) become in the new variables

$$z_x(\xi) = 0, \tag{4.26}$$ $$y(\xi) = -\omega^{-2}. \tag{4.27}$$

The first is a transcendental equation in $\xi$, to be solved for its lowest root, which is found to be $\xi = .98618$.[17] This value inserted in (4.27) determines $\omega$, which is found to be $\omega = .68754$.

Once $\xi$ and $\omega$ are determined, the limit boundary value problem is solved in principle. The function $P(\beta)$ starts with the prescribed value $P(0) = 1$, decreases monotonically, assumes the value zero at $\beta = .941$, and approaches the value $P(\infty) = -\omega^2 = -.47271$ as $\beta \to \infty$, the latter value resulting from (4.17). The function $Q(\beta)$ decreases monotonically and approaches zero as $\beta$ tends to $\infty$. For $Q(0)$ and $P(0)$ we find $Q(0) = -P_\beta(0) = 1.61436$, thus checking (4.15).

We can now discuss the connection[18] between our results from the boundary layer theory and the limit procedure for the interior of the plate described above. The inner edge of the boundary layer is to be identified in the limit with the outer edge $(r \to R)$ of the interior region. Since the value $P(\infty) = -\omega^2$ is the limit value of $p/\bar{p}$ at the inner edge of the boundary layer, $-\omega^2\bar{p}$ is the proper value to be taken for fixing the essential constant for the limit problem in the interior of the plate.[19] Thus we see that the limit membrane stress $p$ in the interior is a tension. The value $Q(\infty) = 0$ is the limit value of $q/\bar{p}$ at the inner edge of the boundary layer and this result is also consistent with the solution $q = 0$ of the equations (4.01).

This solution of the "asymptotic" boundary value problem furnishes limit values for all quantities. From the physical standpoint it is of especial interest to discuss limit values of those quantities which give information on the ultimate stress distribution. The value $Q(0)$ is the limit value of $\eta\bar{q}/\bar{p}$

---

[17] The reason for assuming $z_0 = 4$ was to make $\xi \sim 1$.

[18] A rigorous proof of the validity of the limit process for the *interior* of the plate is given in sec. 10.

[19] This procedure differs from that in Prandtl's boundary layer theory: there the limit state in the interior furnishes a quantity which must be used to determine the solution of the boundary layer equations.

as one sees from the definition (4.04) of $Q$. The value $P_\beta(0)$ is the limit value of $\eta\bar{p}_c/\bar{p}^{3/2} = \bar{p}_c/\bar{p}\lambda$, where $\bar{p}_c = R\bar{p}_r + \bar{p}$ (cf. 1.15) is the circumferential membrane stress at the edge of the plate; this follows for $\lambda \to \infty$ from the identity

(4.28) $$P_c/\bar{p}\lambda = -(1-\beta/\lambda)P_\beta + P/\lambda,$$

where $P$ is defined in (4.04). For the radial bending stress $p_b$ (cf. 1.16) we obtain from (4.04) the formula

(4.29) $$p_b/\bar{p}\lambda = -6\gamma[(1-\beta/\lambda)Q_\beta - (1+\nu)Q/\lambda].$$

The stress $p_b$ attains its maximum at the point where $\beta$ satisfies

$$(1-\beta/\lambda)Q_{\beta\beta} - (2+\nu)Q/\lambda = 0.$$

In the limit ($\lambda \to \infty$) this equation reduces to $Q_{\beta\beta} = 0$; in view of (4.09) it is satisfied at the point $\beta = .941$ where $P(\beta) = 0$. It follows that in the limit the point of maximum bending stress lies in the boundary layer. As a result, we find from (1.28) that the limiting value of $p_{b\,\max}/\bar{p}\lambda$ is $-6\gamma Q_\beta(.941) = 1.1123$.

**Asymptotic development.** We proceed to explain a method of developing our solutions $P$ and $Q$ in the neighborhood of $\lambda = \infty$. This leads to approximate formulas for rather large values of $\lambda$. Although we have not proved the validity of such a development, our numerical results indicate strongly that it is justified.

We confine ourselves to the first step of such a development; it could be carried further. We assume that $P(\beta)$ and $Q(\beta)$ considered as functions of $\kappa = \lambda^{-1}$ possess first derivatives with respect to $\kappa$ at $\kappa = 0$, namely

(4.30) $$\delta P(\beta) = \frac{\partial}{\partial\kappa} P(\beta)\,|^{\kappa=0}, \qquad \delta Q(\beta) = \frac{\partial}{\partial\kappa} Q(\beta)\,|^{\kappa=0}.$$

Differentiation of equations (4.05) and (4.06) with respect to $\kappa$ for $\kappa = 0$ leads to the linear differential equations

(4.31) $$\delta P_{\beta\beta} - Q\delta Q = 3P_\beta,$$

(4.32) $$\delta Q_{\beta\beta} + P\delta Q + Q\delta P = 3Q_\beta$$

for $\delta P$ and $\delta Q$, and the boundary conditions

(4.33) $$\delta P(0) = 0,$$

(4.34) $$\delta Q_\beta(0) = (1+\nu)Q_0, \quad \text{where} \quad Q_0 = Q(0).$$

Differentiation of (4.07) with respect to $\kappa$ yields

$$\delta P_\beta(\kappa^{-1}) - \kappa^{-2}P_{\beta\beta}(\kappa^{-1}) = 0 \quad \text{and} \quad \delta Q_\beta(\kappa^{-1}) - \kappa^{-2}Q_{\beta\beta}(\kappa^{-1}) = 0.$$

Our numerical calculations indicate that the second terms in each of the latter equations tend to zero as $\kappa \to 0$. We assume, then, as boundary conditions

$$(4.35) \qquad \delta P_\beta(\infty) = 0, \qquad\qquad (4.36) \qquad \delta Q_\beta(\infty) = 0;$$

this assumption seems justified by the results. In equations (4.31) to (4.34) the functions $P(\beta)$ and $Q(\beta)$ refer to the solutions of the limit boundary value problem treated earlier in this section.

Before solving the differential equations we note the relation

$$(4.37) \qquad \delta P_\beta + \delta Q = 6/5 \quad \text{for} \quad \beta = 0,$$

which will serve as a useful check on the solution. To prove (4.37) we start with the identity

$$(4.38) \qquad [P_\beta \delta P_\beta - Q_\beta \delta Q_\beta - PQ\delta Q - Q^2 \delta P/2]_\beta = 3(P_\beta{}^2 - Q_\beta{}^2)$$

which follows from (4.31), (4.32), (4.08), (4.09). Integration yields

$$(4.39) \qquad P_\beta \delta P_\beta - Q_\beta \delta Q_\beta - PQ\delta Q - Q^2 \delta P/2 |^0 = -3 \int_0^\infty (P_\beta{}^2 - Q_\beta{}^2)\, d\beta,$$

where the boundary conditions for $\beta = \infty$ have been used. The right member can be evaluated: we multiply (4.08) by $P$, (4.09) by $-Q$, add, and integrate from 0 to $\infty$. The result, after integration by parts, and use of the boundary conditions for $\beta = \infty$, can be written

$$\int_0^\infty (P_\beta{}^2 - Q_\beta{}^2 + 3Q^2P/2)\, d\beta = -PP_\beta + QQ_\beta |^0 = Q(0) = Q_0,$$

where (4.10), (4.11), and (4.15) have been used. In view of (4.14) this yields

$$(4.40) \qquad (5/2) \int_0^\infty (P_\beta{}^2 - Q_\beta{}^2)\, d\beta = Q_0.$$

Insertion of this in (4.39), use of (4.10), (4.11), (4.15), and division by $-Q_0$ gives

$$(4.41) \qquad \delta P_\beta + \delta Q + \tfrac{1}{2} Q \delta P |^0 = 6/5$$

and (4.37) follows from (4.33).

It is possible to obtain solutions of the homogeneous differential equations $(4.31)_0$ and $(4.32)_0$ derived from (4.31) and (4.32). One considers a one-parameter set of solutions $P(\beta, \alpha)$, $Q(\beta, \alpha)$ of the asymptotic differential equations (4.08) and (4.09); the derivatives of $P$ and $Q$ with respect to $\alpha$ satisfy the homogeneous equations $(4.31)_0$ and $(4.32)_0$, as one easily verifies.

One sees readily that $P(\beta+\alpha)$, $Q(\beta+\alpha)$ is such a set; the derivatives of these with respect to $\alpha$ for $\alpha=0$, i. e.,

(4.42) $$\delta P = P_\beta, \qquad \delta Q = Q_\beta$$

constitute, therefore, solutions of $(4.31)_0$ and $(4.32)_0$. We note the following values[20]

(4.43) $$\delta P = -Q_0, \quad \delta Q = 0, \quad \delta P_\beta = Q_0^2/2, \quad \delta Q_\beta = -Q_0 \text{ for } \beta = 0,$$

which follow from (4.10) and (4.15).

A second one-parameter set is formed by $\alpha^2 P(\alpha\beta)$, $\alpha^2 Q(\alpha\beta)$, as can be readily verified; the derivatives of these with respect to $\alpha$ for $\alpha=1$, i. e.,

(4.44) $$\delta P = \beta P_\beta + 2P, \qquad \delta Q = \beta Q_\beta + 2Q$$

form a second solution of $(4.31)_0$ and $(4.32)_0$. We note the following values[20] for this solution:

(4.45) $$P = 2, \quad Q = 2Q_0, \quad P = -3Q_0, \quad Q = 0 \text{ for } \beta = 0,$$

which follow from (4.10) and (4.15).

We proceed now to construct a solution of the non-homogeneous equations (4.31) and (4.32) satisfying the conditions (4.35) and (4.36). Since these conditions are satisfied by the above solutions of the homogeneous equations $(4.31)_0$ and $(4.32)_0$ we may satisfy the remaining boundary conditions by adding an appropriate linear combination of the latter solutions.

We introduce the new variable $x=\xi e^{-\omega\beta}$ as in (4.16) and the new functions $\delta y$ and $\delta z$ by

(4.46) $$\delta P = 3\omega\delta y, \qquad \delta Q = -3\sqrt{2}\omega\delta z.$$

With the notation of (4.17) and (4.18) equations (4.31) and (4.32) become

(4.47) $$x(x\delta y_x)_x + 2z\delta z = xy_x,$$

(4.48) $$x(x\delta z_x)_x - y\delta z - z\delta y = xz_x.$$

Assuming for $\delta y$ and $\delta z$ the power series

(4.49) $$\delta y = \sum_{k=0}^{\infty} (-1)^k \delta y_k x^{2k},$$

(4.50) $$\delta z = \sum_{m=0}^{\infty} (-1)^m \delta z_m x^{2m+1},$$

and the series (4.21) and (4.22) for $y$ and $z$ we obtain by insertion in (4.47) and (4.48) the relation

---

[20] We observe that the left member of (4.41) vanishes, as it should, for these values.

(4.51) $$(2k)^2 y_k = \sum_{n+m=k-1} z_m \delta z_n + k y_k, \qquad (k = 1, 2, \cdots).$$

(4.52) $$(2m+1)^2 \delta z_m = \sum_{k+n=m} (y_k \delta z_n + z_n \delta y_k) + (2m+1) z_m, \qquad (m = 0, 1, \cdots).$$

The coefficient of $\delta z_m$ in the right member of (4.52) is $y_0 = 1$; hence we may replace (4.52) by

(4.53) $$2m(2m+2)\delta z_m = \sum_{n=0}^{m-1} y_{m-n} \delta z_n + \sum_{k+m=n} z_n \delta y_k + (2m+1) z_m, \qquad (m = 0, 1, 2, \cdots).$$

This and (4.51) constitute proper recursion formulas for the series (4.49) and (4.50). The coefficient $\delta y_0$ is fixed by (4.53) for $m = 0$; its value is $-1$. The coefficient $\delta z_0$ can be chosen arbitrarily; it is convenient to take $\delta z_0 = 0$. The values $y_k$ and $z_m$ have been previously calculated from (4.23) and (4.24); the coefficients $\delta y_k$ and $\delta z_m$ can then be determined from (4.51) and (4.53). The convergence of the resulting series for $\delta y$ and $\delta z$ appears to be very satisfactory.

In this way we find a particular solution $\delta P$, $\delta Q$ of the non-homogeneous equations (4.31), (4.32); by adding a proper linear combination of the two previously found solutions of $(4.31)_0$ and $(4.32)_0$ we obtain a solution of (4.31) and (4.32) satisfying the conditions (4.33), (4.34). We give a few numerical results:[21]

$$\delta P = 0, \quad \delta Q = -.74, \quad \delta P_\beta = 1.94, \quad \delta Q_\beta = 2.13 \text{ for } \beta = 0,$$
$$\delta P = -1.03, \quad \delta Q = 0, \quad \delta P_\beta = 0, \quad \delta Q_\beta = 0 \text{ for } \beta = \infty.$$

Hence we have the following approximation formulas for large $\lambda$ (that is, for large $N$): for $\beta = 0$

$$P = 1, \quad Q = 1.61 - .74/\lambda, \quad P_\beta = -1.61 + 1.94/\lambda, \quad Q = 2.13/\lambda;$$

for $\beta = \infty$ we have

$$P = -.47 - 1.03/\lambda, \quad Q = 0, \quad P_\beta = 0, \quad Q_\beta = 0.$$

It is clear that the preceding calculations yield also approximate formulas for all physical quantities such as stresses and deflections.

## PART II.

In Part II of this paper we consider the purely mathematical aspects of the boundary value problems which were discussed in Part I from the point of view of explicit numerical solution.

---

[21] Note that our results check with formula (4.37).

**5. Minimum and boundary value problems for the finite case.** Our subsequent discussion will be based upon the possibility of formulating minimum problems equivalent to our boundary value problems.[22] In this section we consider the problems for finite $\lambda$. The functional to be minimized, essentially the total potential energy,[23] is

$$(5.01)\qquad V=\eta^2\int_0^R q_r^2(r)r^3dr+(1+\nu)R^2q^2(R)\\ -\bar{p}R^{-2}\int_0^R q^2(r)r^3dr+\int_0^R p_r^2(r)r^3dr.$$

We do not vary $p$ and $q$ independently. Rather, $q$ must be varied while $p$ is considered a functional in $q$ through the differential equation (1.08) and the boundary conditions (1.10) and (1.12).

It is convenient to work with the invariant quantities $p^*$, $q^*$, and $r_*$ defined at the end of sec. 1, but we omit the * from now on. We introduce the functionals

$$(5.02)\qquad qHq=\int_0^1 q^2(r)r^3dr,$$

$$(5.03)\qquad qDq=\int_0^1 q_r^2(r)r^3dr+(1+\nu)q^2(1),$$

$$(5.04)\qquad p_rKp_r=\int_0^1 p_r^2(r)r^3dr,$$

and

$$(5.05)\qquad V_\lambda[q]=qDq-\lambda^2qHq+p_rKp_r.$$

The new functional $V_\lambda$ to be minimized is related to (5.01) through $V=\eta^4R^2V_\lambda$.

We define $p(r)$ and $p_r(r)$ as functionals in $q(r)$ by means of the formulas

$$(5.06)\qquad p_r(r)=\tfrac{1}{2}r^{-3}\int_0^r q^2(r_1)r_1^3dr_1.$$

$$(5.07)\qquad p(r)=\lambda^2-\int_r^1 p_r(r_1)dr_1.$$

We are now in a position to formulate *admissibility conditions.* By admissible functions we mean functions $q(r)$ continuous in $0<r\leq 1$ which possess $L^2$-integrable[24] derivatives in every interval $0<\epsilon\leq r\leq 1$ and for

---

[22] Incidental use of this fact was made in sec. 2.

[23] The total potential energy is given by

$$\pi hE[V-(1-\nu)R^2\bar{p}^2].$$

[24] For a method avoiding Lebesgue derivatives which could be applied here, see [8].

which the integrals $H$, $D$, and $K$ are finite, $p_r$ being considered a functional in $q(r)$ through (5.06).

The *minimum problem* $M_\lambda$ is that of minimizing $V_\lambda[q]$ with respect to admissible functions $q(r)$.

The first variation $\delta V_\lambda$ of $V_\lambda$ is

$$\delta V_\lambda[q] = qD\delta q - \lambda^2 qH\delta q + p_r K\delta p_r, \tag{5.08}$$

where the bilinear forms corresponding to (5.02), (5.03), and (5.04) have been used; $\delta q$ is any admissible function and

$$\delta p_r = r^{-3}\int_0^r q\delta q r_1^3 dr_1. \tag{5.09}$$

$V_\lambda$ is said to be stationary for a function $q$ if $\delta V_\lambda[q] = 0$, for all admissible $\delta q$. We refer to the problem of making $V_\lambda$ stationary in this sense as the problem $S_\lambda$. A solution of $M_\lambda$ is also a solution of $S_\lambda$; this conclusion results from the standard reasoning.

We formulate the *boundary value problem* $B_\lambda$ as follows; an admissible function $q$ is to be found which possesses a continuous second derivative and satisfies the differential equation

$$Gq + pq = 0, \qquad G = r^{-3}\frac{d}{dr}r^3\frac{d}{dr}, \tag{5.10}$$

and the boundary condition

$$B_\nu q = rq_r + (1+\nu)q = 0 \quad \text{for} \quad r = 1. \tag{5.11}$$

The function $p$ in (5.10) is defined by (5.07) and (5.06). These conditions imply that $p$ satisfies the differential equation

$$Gp = q^2/2, \tag{5.12}$$

and the boundary condition

$$p = \lambda^2 \quad \text{for} \quad r = 1. \tag{5.13}$$

The conditions $p_r = q_r = 0$ at $r = 0$ which were used in calculating the solutions in sec. 1 are not required; we have sufficiently characterized the behavior of $p$ and $q$ at $r = 0$ through the admissibility conditions[25] that the integrals $H$, $D$, and $K$ be finite.

The problems $S_\lambda$ and $B_\lambda$ are related as follows: 1) A solution of $B_\lambda$ solves $S_\lambda$. 2) If an admissible function $q$ solves $S_\lambda$ it possesses a continuous second derivative and solves $B_\lambda$. These statements are proved in sec. 7 (cf.

[25] It could be shown that a solution of the boundary value problem which satisfies the admissibility conditions would also satisfy the regularity conditions at $r = 0$.

Theorems 7.1 and 7.2) in terms of convenient new variables, which at the same time permit proofs of the corresponding statements for the asymptotic case. We may mention here that the connection between the two problems is based on the Green's formula

$$\delta V_\lambda(q) = -\int_0^1 (Gq + pq)\delta q r^3 dr + B_\nu q \delta q|^1, \tag{5.14}$$

which, in terms of new variabes, is justified in sec. 7 (cf. formula 7.21).

**6. Multiplicity of solutions of the boundary value problem.** In this section we are interested in the *number* of possible *distinct* solutions $q$ of the boundary value problem $B_\lambda$ in their dependence upon the parameter $\lambda$. We proceed first to give a general description of what can be expected to occur with increasing $\lambda$. For sufficiently small $\lambda$ there will be only one solution, i. e., $q \equiv 0$. This holds up to $\lambda = \lambda_0$, where $\lambda_0$ is the lowest eigenvalue of the linearized buckling problem. Beyond $\lambda_0$ two new solutions will appear which differ only in sign. We shall prove in this section that these solutions together with $q \equiv 0$ are the only solutions as long as $\lambda < \lambda_1$, where $\lambda_1$ is the second eigenvalue of the linear buckling problem. It could be proved, for example by the perturbation method, that two additional solutions will appear when $\lambda$ passes the next critical value $\lambda_1$. It seems likely that there would be exactly $2n+1$ solutions when $\lambda_{n-1} < \lambda < \lambda_n$, one the solution $q \equiv 0$, and $n$ other pairs differing only in sign. In this section we prove less, namely:

THEOREM 6.1. *For $0 \leq \lambda \leq \lambda_0$ the sole solution of $B_\lambda$ is $q = 0$.*

THEOREM 6.2.[26] *For $\lambda_0 < \lambda < \lambda_1$ there are at most three solutions of $B_\lambda$; one of these is $q \equiv 0$ and the other two differ only in sign.*

We shall make use of the following properties[27] of the eigenvalues $\lambda_0^2$ and $\lambda_1^2$. The inequality

$$qD_\lambda q \equiv qDq - \lambda^2 qHq > 0 \tag{6.1}$$

holds for admissible functions $q \not\equiv 0$ (cf. sec. 5) if $\lambda < \lambda_0$, equality holding for the eigenfunction $q \equiv q_0$; the inequality (6.1) holds also for $\lambda_0 < \lambda < \lambda_1$ if the additional restriction

---

[26] This theorem could be easily generalized. The essential properties used are readily seen to be: $D + H$ is a positive definite quadratic form, $H$ is a completely continuous quadratic form with respect to $D + H$ as unit form, and $K$ is a positive definite, centro-symmetric, and convex functional of degree greater than one. The reasoning given here could, with slight adaptations, be taken over for such a generalization.

[27] Cf. Courant-Hilbert [3] Volume I, Chapter VI.

$$(6.2) \qquad qHq_0 = 0, \qquad q \not\equiv 0,$$

is imposed, where $q_0$ is the eigenfunction belonging to $\lambda_0$.

In sec. 5 it was stated that solutions $q$ of the boundary value problem $B_\lambda$ also solve $S_\lambda$, i. e., they are completely characterized by the statement that $\delta V_\lambda[q] = 0$ for all admissible $\delta q$.

If $q$ renders $V_\lambda[q]$ stationary, the identity

$$(6.3) \qquad qD_\lambda q + 2p_rKp_r = 0$$

holds. It results when $\delta q$ is replaced by $q$ in (5.08) and (5.09).

For $\lambda \leq \lambda_0$ we may apply (6.1), and (6.3) leads to $p_rKp_r = 0$. In view of (5.04) and (5.06) this is possible only for $q \equiv 0$. Thus Theorem 6.1 is proved.

In the proof of Theorem 6.2 we make use of the fact that the positive definite form $p_rKp_r$ is a *convex* functional (of fourth degree) in $q$. By convexity we mean that the second variation $\delta p_rK\delta p_r + \delta^2 p_rKp_r$ of $p_rKp_r$ is positive. This is the case since the second term is positive because $p_r > 0$ and $\delta^2 p_r = r^{-3}\int_0^r (\delta q)^2 r_1^3 dr_1 > 0$, while the first term is obviously positive.

The proof of Theorem 6.2 will be given indirectly. Assume, then, that there are *two* linearly independent solutions $q^{(1)} \not\equiv 0$, $q^{(2)} \not\equiv 0$. We introduce the linear combinations $q = \alpha_1 q^{(1)} + \alpha_2 q^{(2)}$ and consider the homogeneous forms

$$Q(\alpha_1, \alpha_2) = qD_\lambda q, \qquad P(\alpha_1, \alpha_2) = p_rKp_r.$$

Since $q^{(1)}$ and $q^{(2)}$ make $V_\lambda$ stationary, the sum $V_\lambda = Q(\alpha_1, \alpha_2) + P(\alpha_1, \alpha_2)$ is stationary for $(1,0)$, $(-1,0)$ and for $(0,1)$, $(0,-1)$. Our Theorem 6.2 is proved if we can show that all stationary points of $V_\lambda = P + Q$ lie on the same straight line through the origin in the $(\alpha_1, \alpha_2)$-plane, in contradiction with the assumption that $(1,0)$ and $(0,1)$ are stationary points.

The quadratic form $Q(\alpha_1, \alpha_2)$ is indefinite since 1. $Q(1,0) = q^{(1)}D_\lambda q^{(1)} < 0$ as we see from (6.3); 2. the linear set $q$ contains a function $\hat{q} = \hat{\alpha}_1 q^{(1)} + \hat{\alpha}_2 q^{(2)} \not\equiv 0$ satisfying $\hat{q}Hq_0 = 0$ and since $\lambda < \lambda_1$ we know that (6.1) holds for $\hat{q}$; thus $Q(\hat{\alpha}_1, \hat{\alpha}_2) = \hat{q}D_\lambda\hat{q} > 0$. The form $P(\alpha_1, \alpha_2)$ is of the fourth degree, and is positive definite and convex since, as we have seen, this is the case for $p_rKp_r$. Both forms are clearly centro-symmetric and stationary points occur in symmetrically located pairs.

Let $\bar{\alpha}_1, \bar{\alpha}_2$ be any values of $\alpha_1$ and $\alpha_2$ $(\alpha_1^2 + \alpha_2^2 \not\equiv 0)$ for which $V_\lambda = Q(\alpha_1, \alpha_2) + P(\alpha_1, \alpha_2)$ is stationary, i. e., for which $\delta V_\lambda = \delta Q + \delta P = 0$. As a consequence we have

$$(6.4) \qquad -Q_{\alpha_1}d\alpha_1 - Q_{\alpha_2}d\alpha_2 = P_{\alpha_1}d\alpha_1 + P_{\alpha_2}d\alpha_2$$

for $\alpha_1 = \bar{\alpha}_1$, $\alpha_2 = \bar{\alpha}_2$. Let $\bar{P}$ and $\bar{Q}$ be the values of $P$ and $Q$ for $\alpha_1 = \bar{\alpha}_1$, $\alpha_2 = \bar{\alpha}_2$. In view of (6.4) the curves $P(\alpha_1, \alpha_2) = \bar{P}$ and $Q(\alpha_1, \alpha_2) = \bar{Q}$ have a common tangent at $(\bar{\alpha}_1, \bar{\alpha}_2)$.

Consider a second point where $V_\lambda = P + Q$ is stationary and where, consequently, two curves $P = \text{const.}$, $Q = \text{const.}$ have a common tangent. By a similarity transformation in the $(\alpha_1, \alpha_2)$-plane we can transform $P = \text{const.}$ into the original curve $P = \bar{P}$ while the equation for the new curve obtained from $Q = \text{const.}$ by the transformation may be written in the form $Q = \epsilon\bar{Q}$ with $\epsilon > 0$. In order to show that such a second stationary point is on the same ray as $(\bar{\alpha}_1, \bar{\alpha}_2)$ it is obviously sufficient to show that it concides with $(\bar{\alpha}_1, \bar{\alpha}_2)$ after the similarity transformation. This we accomplish by proving that the set of curves $Q = \epsilon\bar{Q}$ contains only one curve, namely $Q = \bar{Q}$, tangent to $P = \bar{P}$ and that $Q = \bar{Q}$ has only two symmetrically located points of contact with $P = \bar{P}$.

The set of points $P \leq \bar{P}$ is convex since its boundary is a level line of a convex surface. The curves $Q = \epsilon\bar{Q}$ constitute a set of hyperbolas since $Q(\alpha_1, \alpha_2)$ is an indefinite quadratic form; we need evidently consider only one branch of these hyperbolas. Each such branch is the boundary of an unbounded convex set $S_\epsilon$. In addition it is to be noted that the origin is not contained in $S_\epsilon$, but is an inner point of the set $P \leq \bar{P}$.

When the curves $P = \bar{P}$ and $Q = \epsilon\bar{Q}$ have a common tangent, the two convex point sets $P \leq \bar{P}$ and $S_\epsilon$ lie entirely on the same side of the tangent line, or they lie entirely on opposite sides. The first case is excluded since the tangent to the hyperbola $Q = \epsilon\bar{Q}$ separates the set $S_\epsilon$ from the origin and this point lies in the interior of the set $P \leq \bar{P}$. In the second case the two convex sets have only one point in common since $S_\epsilon$ is bounded by a hyperbola and $P \leq \bar{P}$ is convex.

We have assumed that a common tangent line exists for $\epsilon = 1$; such a tangent line separates the sets $P \leq \bar{P}$ and $S_1$. For every $\epsilon > 1$ it is obvious that the sets $P \leq \bar{P}$ and $S_\epsilon$ have no common points. For every $\epsilon < 1$ it is obvious that the two sets have inner points in common; hence a possible new tangency would fall under the first case which has already been excluded. With this we have established the contradiction which proves our theorem.

It appears that the type of reasoning used above would not be sufficient to give a corresponding theorem for $\lambda > \lambda_1$.

**7. Simultaneous formulation of finite and asymptotic problems.** For the investigations of sections 8 and 9 it is convenient and useful to formulate our problems in terms of certain new variables. At the same time the new formulations make possible simultaneous treatment of the finite and the asymptotic problems.

The new variables are (cf. sec. 5)

(7.01) $P = p/\bar{p}$ and (7.02) $Q = q/\bar{p}$;

the new independent variable is

(7.03) $\beta = \lambda(r^{-2} - 1)/2$ with the domain $0 \leq \beta < \infty$.

With the notation

(7.04) $\kappa = \lambda^{-1}$ we have[28] (7.05) $r = (1 + 2\kappa\beta)^{-1/2}$.

The parameter $\kappa$ occurs essentially only in the combination

(7.06) $$\rho^\kappa(\beta) = (1 + 2\kappa\beta)^{-6}.$$

In the following formulations we admit $\kappa \geq 0$; the cases $\kappa > 0$ are equivalent to the cases for finite $\lambda$ (cf. sec. 5), while the case $\kappa = 0$ corresponds to the transition $\lambda \to \infty$ and will be seen to represent the asymptotic case.

We introduce the forms

(7.07) $$H^\kappa[Q] = \int_0^\infty Q^2(\beta)\rho^\kappa(\beta)\,d\beta,$$

(7.08) $$D^\kappa[Q] = \int_0^\infty Q_\beta^2(\beta)\,d\beta + \kappa(1+\nu)Q^2(0),$$

and

(7.09) $$K^\kappa[Q] = \int_0^\infty P_\beta^2(\beta)\,d\beta,$$

where $P_\beta$ is a functional in $Q$ through

(7.10) $$P_\beta = P_\beta^\kappa[Q] = -(1/2)\int_\beta^\infty Q^2(\beta_1)\rho^\kappa(\beta_1)\,d\beta_1.$$

The functional[29] to be minimized is

(7.11) $$W^\kappa[Q] = D^\kappa[Q] - H^\kappa[Q] + K^\kappa[Q].$$

By *admissible functions* we mean functions $Q(\beta)$ continuous in $0 \leq \beta < \infty$ with $L^2$-integrable derivatives in every interval $0 \leq \beta \leq b < \infty$ and for which the integrals in (7.07), (7.08), and (7.09) are finite, $P_\beta$ being defined by (7.10).

The *minimum problem* $M^\kappa$ is that of minimizing $W^\kappa[Q]$, the problem $S^\kappa$ is that of making $W^\kappa[Q]$ stationary, in each case with respect to admissible functions $Q$.

---

[28] The variable $\beta$ here is not the same as that of sec. 4 but it yields the same transition to the asymptotic case. The transformation $r = (1 - \kappa\beta)$ of sec. 4 could be obtained as the term of first order in the development of (7.05) with respect to $\kappa$.

[29] The relation between $V_\lambda$ (see (5.5)) and $W^\kappa$ is $W^\kappa = \kappa^5 V_\lambda$.

The boundary value problem $B^\kappa$ requires the determination of an admissible function $Q(\beta)$ possessing a continuous second derivative and satisfying the differential equation

$$Q_{\beta\beta} + \rho^\kappa P Q = 0 \tag{7.12}$$

and the boundary condition

$$Q_\beta(0) - \kappa(1+\nu)Q(0) = 0. \tag{7.13}$$

The function $P(\beta)$ in (7.12) is defined by

$$P(\beta) = 1 + \int_0^\infty P_\beta(\beta_1)\,d\beta_1, \tag{7.14}$$

where $P_\beta$ is given by (7.10). The function $P$ therefore satisfies the differential equation

$$P_{\beta\beta} = \rho^\kappa Q^2/2 \tag{7.15}$$

and the boundary condition

$$P(0) = 1. \tag{7.16}$$

The problems $M^\kappa$, $S^\kappa$, and $B^\kappa$ ($\kappa > 0$) are the equivalents, in the new variables, of the problems $M_\lambda$, $S_\lambda$, and $B_\lambda$ of sec. 5.

To obtain the formulation of the asymptotic problems $M^0$, $S^0$, and $B^0$ we need only set $\kappa = 0$ in the preceding. Since $\kappa = 0$ implies $\rho^0 = 1$ (cf. (7.06)), this formulation yields the same differential equations and boundary conditions as in sec. 4 except that the conditions used in sec. 4 at $\beta = \infty$ are here replaced by the admissibility conditions.

The connection between the problems $S^\kappa$ and $B^\kappa$ (including $\kappa = 0$) is based on two "Green's" formulas. They refer to the first variation

$$\delta W^\kappa[Q] = \int_0^\infty [Q_\beta \delta Q_\beta - \rho^\kappa Q \delta Q + P_\beta \delta P_\beta]\,d\beta + \kappa(1+\nu)Q(0)\delta Q(0), \tag{7.17}$$

where $\delta P_\beta$ is defined, in accordance with (7.10), by

$$\delta P_\beta = -\int_\beta^\infty Q(\beta_1)\delta Q(\beta_1)\rho^\kappa(\beta_1)\,d\beta_1. \tag{7.18}$$

It is convenient to state first the following lemma, the proof of which will be given later:

LEMMA 7.1. *If $f(\beta)$ and $g(\beta)$ are continuous functions with $L^2$-integrable derivatives for which $\int_0^\infty f^2 d\beta < \infty$, $\int_0^\infty g^2 d\beta < \infty$, then there is a special sequence $\beta \to \infty$ on which*

$$f(\beta)g_\beta(\beta) \to 0. \tag{7.19}$$

COROLLARY. *If $f_1, g_1$ and $f_2, g_2$ are pairs of functions with the properties of $f$ and $g$ in Lemma 7.1 there exists a special sequence $\beta \to \infty$ on which (7.19) holds for both pairs.*

We consider (for finite $b > 0$) the relation

$$\int_0^b P_\beta \delta P_\beta d\beta = -\int_0^b (P-1)\delta P_{\beta\beta} d\beta + [(P-1)\delta P_\beta]_0^b,$$

obtained by product integration. Lemma 7.1 with $f = (P-1)$ and $g_\beta = \delta P_\beta$ [30] shows that $[(P-1)\delta P_\beta]^b \to 0$ if $b \to \infty$ on a special sequence. Hence we have

$$\int_0^\infty P_\beta \delta P_\beta d\beta = -\int_0^\infty (P-1)\delta P_{\beta\beta} d\beta + [(P-1)\delta P_\beta]_0.$$

If this and the relations $P(0) = 1$, $\delta P_{\beta\beta} = Q\delta Q$ (cf. 7.18) are used one obtains from (7.17) the first Green's formula:

$$(7.20) \qquad \delta W^\kappa[Q] = \int_0^\infty [Q_\beta \delta Q_\beta - \rho^\kappa P Q \delta Q] d\beta + \kappa(1+\nu)Q(0)\delta Q(0).$$

It holds for all admissible $Q$ and $\delta Q$ and $\int_0^\infty$ means $\lim \int_0^b$ when $b \to \infty$ on a special sequence.

If $Q$ possesses a second derivative we may write

$$\int_0^b Q_\beta \delta Q_\beta d\beta = -\int_0^b Q_{\beta\beta} \delta Q_\beta d\beta + [Q_\beta \delta Q]_0^b.$$

The corollary to Lemma 7.1 applied to $[(P-1)\delta P_\beta]^b$ and $[Q_\beta \delta Q]^b$ yields the existence of a common special sequence $b \to \infty$ on which both expressions tend to zero. If this sequence is used and if $\int_0^\infty Q_\beta \delta Q_\beta d\beta$ is replaced by $-\int_0^\infty Q_{\beta\beta} \delta Q_\beta d\beta + [Q_\beta \delta Q]_0$ in (7.20) we obtain the second Green's formula

$$(7.21) \qquad \delta W^\kappa[Q] = -\int_0^\infty [Q_{\beta\beta} + \rho^\kappa P Q]\delta Q d\beta - [Q_\beta(0) + \kappa(1+\nu)Q(0)]\delta Q(0),$$

which holds for admissible functions $Q$ possessing second derivatives, the integral again being "conditional." Formula (7.21) yields immediately

[30] $\int_0^\infty \delta P_\beta{}^2 d\beta < \infty$ follows from the admissibility conditions on $Q$ and $\delta Q$ and the Schwarz inequality applied to (7.18).

THEOREM 7.1.[31] *A solution of $B^\kappa$ solves $S^\kappa$.*

The converse also holds:

THEOREM 7.2.[31] *A solution of $S^\kappa$ (hence also a solution of $M^\kappa$) possesses a continuous second derivative and solves $B^\kappa$.*

To prove Theorem 7.2 we make use of the following well known

LEMMA 7.2.[32] *Let $R(\beta)$ be an $L^2$-integrable function and $S(\beta)$ be a continuous function such that $\int_0^\infty RT_\beta d\beta = \int_0^\infty ST d\beta$ holds for all continuous functions $T(\beta)$ with $L^2$-integrable derivatives $T_\beta$ which vanish identically in the neighborhood of $\beta = 0$ and of $\beta = \infty$. Then $R(\beta)$ coincides (almost everywhere) with a function $\bar{R}$ which possesses the continuous derivative $-S$. We note in addition that $\bar{R} = R$ in case $R$ is the derivative of a continuous function.*

We apply this lemma to $R = Q_\beta$, $S = \rho^\kappa PQ$, where $Q$ is a solution of $S^\kappa$. Since $\delta W^\kappa[Q] = 0$ for the admissible variations $\delta Q = T$ it follows from (7.20) that $Q$ possesses the continuous second derivative $-\rho^\kappa PQ$; thus $Q$ satisfies the differential equation (7.12). We may now apply (7.21); it yields $0 = [Q_\beta(0) + \kappa(1+\nu)Q(0)]\delta Q(0)$ and since $\delta Q(0)$ is arbitrary the boundary condition (7.13) is satisfied. Hence Theorem 7.2 is proved.

We have to supply the proof of Lemma 7.1 and its corollary. This will be done with the aid of two additional lemmas of which we shall make frequent use in later sections.

LEMMA 7.3. *"Jacobi's identity." If $f(\beta)$ is a function with $L^2$-integrable derivative and the positive function $\omega(\beta)$ possesses continuous second derivatives, then*

$$(7.22) \qquad \int_a^b [f_\beta^2 + \omega^{-1}\omega_{\beta\beta}f^2]\,d\beta = \int_a^b \omega^2[(\omega^{-1}f)_\beta]^2 d\beta + \omega^{-1}\omega_\beta f^2\Big|_a^b.$$

The identity (7.22) follows from

$$\begin{aligned}\int_a^b \omega^2[(\omega^{-1}f)_\beta]^2 d\beta &= \int_a^b (f_\beta^2 - 2\omega^{-1}\omega_\beta f f_\beta + \omega^{-2}\omega_\beta^2 f^2)\,d\beta \\ &= \int_a^b [f_\beta^2 - (\omega^{-1}\omega_\beta f^2)_\beta + \omega^{-1}\omega_{\beta\beta}f^2]\,d\beta \\ &= \int_a^b (f_\beta^2 + \omega^{-1}\omega_{\beta\beta}f^2)\,d\beta - \omega^{-1}\omega_\beta f^2\Big|_a^b.\end{aligned}$$

---

[31] These theorems are equivalent (for $\kappa > 0$) to the statements made at the end of sec. 5.

[32] The lemma can be proved in much the same way as the lemma of du Bois-Reymond to which it reduces if $R(\beta)$ is assumed continuous.

LEMMA 7.4. *If $f(\beta)$ is a continuous function with $L^2$-integrable derivative for which* $\int_0^\infty f^2(\beta)d\beta < \infty$, *then*

$$\beta^{-1/2}f(\beta) \to 0 \text{ as } \beta \to \infty. \tag{7.23}$$

To prove the lemma we start with Jacobi's identity for $\omega = \beta^{1/2}$:

$$\int_a^b [f_\beta{}^2 - \tfrac{1}{4}\beta^{-2}f^2]d\beta = \int_a^b \beta(\beta^{-1/2}f)_\beta{}^2 d\beta + \tfrac{1}{2}[\beta^{-1}f^2]_a^b,$$

or

$$\tfrac{1}{4}\int_a^b \beta^{-2}f^2 d\beta + \int_a^b \beta(\beta^{-1/2}f)_\beta{}^2 d\beta + \tfrac{1}{2}b^{-1}f^2(b) = \int_a^b f_\beta{}^2 d\beta + \tfrac{1}{2}a^{-1}f^2(a).$$

For $b \to \infty$ the right member is bounded; hence the terms of the left member are bounded since all are positive. From

$$\int_a^\infty \beta^{-1}f^2 \frac{d\beta}{\beta} = \int_a^\infty \beta^{-2}f^2 d\beta < \infty$$

we see that $\beta^{-1}f^2(\beta) \to 0$ at least on a special sequence $\beta \to 0$. If $b$ is taken on that special sequence we obtain

$$\tfrac{1}{4}\int_a^\infty \beta^{-2}f^2 d\beta + \int_a^\infty (\beta^{-1/2}f)^2 d\beta = \int_a^\infty f_\beta{}^2 d\beta + \tfrac{1}{2}a^{-1}f^2(a),$$

which, for $a \to \infty$, yields $a^{-1}f^2(a) \to 0$ and the Lemma 7.4 is proved.

We turn now to the proof of Lemma 7.1. From $\int_0^\infty \beta g_\beta{}^2\beta^{-1}d\beta < \infty$ we infer that a special sequence exists on which $\beta^{1/2}g_\beta \to 0$; in view of Lemma 7.4 we have $\beta^{-1/2}f(\beta) \to 0$ and (7.19) follows immediately.

For the proof of the corollary it is sufficient to observe that from $\int_0^\infty (\beta[g_{1\beta}]^2 + \beta[g_{2\beta}]^2)(\beta^{-1}d\beta) < \infty$ a special sequence $\beta \to \infty$ exists on which $\beta[g_{1\beta}]^2 + \beta[g_{2\beta}]^2 \to 0$ and that therefore also $\beta^{1/2}g_{1\beta} \to 0$ and $\beta^{1/2}g_{2\beta} \to 0$.

**8. Uniqueness theorems.** In sec. 6 we proved that the boundary value problem $B_\lambda$ has, apart from sign, only one solution $q \not\equiv 0$ for a certain range of values of $\lambda$ ($\lambda < \lambda_1$). It was also indicated there that the problem $B_\lambda$ will possess more and more distinct solutions as $\lambda$ increases beyond $\lambda_1$. The situation is quite different as regards the minimum problem $M_\lambda$. We shall prove in this section (with the notation of sec. 7):

THEOREM 8.1. *There is at most one solution $Q$ of the minimum problem $M^\kappa$, apart from the sign of $Q$.*

We first dispose of the case in which $Q \equiv 0$ is a solution of $M^\kappa$. In this case the functional $W^\kappa$ is non-negative; otherwise there would be a function $Q$ with $W^\kappa[Q] < 0$ and $W^\kappa[Q] = 0$ would not be the minimum. If $Q$ is *any* solution of $M^\kappa$ with

$$W^\kappa[Q] = D^\kappa[Q] - H^\kappa[Q] + K^\kappa[Q] = 0,$$

then, for constant $\epsilon$

$$W^\kappa[\epsilon Q] = \epsilon^2\{D[Q] - H[Q]\} + \epsilon^4 K[Q] = (\epsilon^4 - \epsilon^2)K[Q]$$

would be negative for $\epsilon < 1$ unless $K[Q] = 0$, which implies $P_\beta = 0$, and $Q^2 = 2P_{\beta\beta} = 0$. Hence $Q \equiv 0$ is the only solution of $M^\kappa$ if $W^\kappa$ is non-negative. From now on we may thus leave aside the case in which $Q \equiv 0$ solves $M^\kappa$.

We have noted in the preceding section that a solution of the minimum problem $M^\kappa$ also solves $S^\kappa$ and the boundary value problem $B^\kappa$ (cf. Theorem 7.2). Hence Theorem 8.1 (for $Q \not\equiv 0$) results from the following two theorems:

THEOREM 8.2. *A solution $Q(\beta) \not\equiv 0$ of $M^\kappa$ is nowhere zero in the interval $0 \leq \beta < \infty$.*

THEOREM 8.3. *A solution $Q(\beta)$ of $B^\kappa$ which is nowhere zero in the interval $0 \leq \beta < \infty$ is, apart from the sign of $Q$, the sole solution of $M^\kappa$.*

An immediate consequence of Theorem 8.3 is the following

COROLLARY.[33] *The problem $B^\kappa$ has only one solution, apart from the sign of $Q$, which is nowhere zero in the interval $0 \leq \beta < \infty$.*

We prove Theorem 8.2 indirectly. Assuming then that the solution $Q(\beta)$ of $M^\kappa$ vanishes for some value $\beta = \beta_0$. We distinguish two cases:

$$1.\ P(\beta) \leq 0 \quad \text{for} \quad \beta \geq \beta_0, \quad \text{and} \quad 2.\ P(\beta) > 0 \quad \text{for} \quad \beta \leq \beta_0.$$

That Case 1 or Case 2 occurs follows from the fact that $P(\beta)$ decreases monotonically from the value $P(0) = 1$.

*Case* 1: We make use of the fact that $Q(\beta)$ solves $B^\kappa$ and hence satisfies the differential equation $Q_{\beta\beta} + \rho^\kappa PQ = 0$. If a solution $Q(\beta)$ of such a differential equation vanishes at a point $\beta = \beta_0$, then, as is well known,

[33] This corollary shows that the solutions for the finite and the asymptotic cases, as treated in sec. 3 and sec. 4, are the sole solutions of $B^\kappa$ provided that the power series used there converge and the transcendental equations are solved. The admissibility conditions are satisfied if the series converge and the non-vanishing of the solutions is guaranteed by taking the lowest root of the transcendental equations, as one easily shows.

$Q_\beta(\beta_0) \neq 0$ unless $Q(\beta) \equiv 0$. Hence, in our case, $Q_\beta(\beta_0) \neq 0$. We may assume $Q_\beta(\beta_0) > 0$. Let $b$ be the greatest value such that $Q(\beta) \geq 0$ for $\beta_0 \leq \beta < b$. In this range we then have $\rho^\kappa PQ \leq 0$, $Q_{\beta\beta} \geq 0$ and, consequently, $Q_\beta(\beta) \geq Q_\beta(\beta_0)$ and $Q(\beta) \geq (\beta - \beta_0) Q_\beta(\beta_0)$. If $b$ were finite the latter inequality would yield $Q(b) > 0$, in contradiction with the definition of $b$. Hence $b = \infty$. The inequality $Q_\beta(\beta) \geq Q_\beta(\beta_0) > 0$ for $\beta_0 \leq \beta < \infty$ is, however, in contradiction with $\int_0^\infty Q^2 d\beta < \infty$. This rules out Case 1.

*Case* 2: We shall dispose of this case by constructing an admissible function $Q^\epsilon$ for which $W[Q^\epsilon] < W[Q]$ in contradiction with the minimum property of $Q$. We first replace $Q$ by $|Q|$ and observe that $W[|Q|] = W[Q]$. We choose $\beta_1 \leq \beta_0$ and $\beta_2 > \beta_0$ such that

$$|Q|_\beta = -|Q_\beta| \text{ for } \beta_1 \leq \beta < \beta_0, \quad |Q|_\beta = |Q_\beta| \text{ for } \beta_0 < \beta \leq \beta_2,$$
$$P > 0 \text{ for } \beta \leq \beta_2.$$

We then introduce a non-negative function $\eta(\beta) \not\equiv 0$ with continuous derivative which has the following properties:

$$\eta(\beta) \geq 0, \ \eta(\beta) \equiv 0 \quad \text{for} \quad \beta \geq \beta_2 \quad \text{and} \quad \beta \leq \beta_1,$$
$$\eta_\beta(\beta) \geq 0 \quad \text{for} \quad \beta \leq \beta_0, \quad \leq 0 \quad \text{for} \quad \beta \geq \beta_0.$$

Upon introduction of the function $Q^\epsilon = |Q| + \epsilon\eta$, we notice that

$$\int_\beta^\infty [Q^\epsilon(\beta_1)]^2 d\beta_1 - \int_\beta^\infty Q^2(\beta_1)\, d\beta_1 = \epsilon\delta^\epsilon(\beta),$$

where

$$\delta^\epsilon(\beta) = 2\int_\beta^\infty \eta(\beta_1)\,|\,Q(\beta_1)\,|\,d\beta_1 + \epsilon\int_\beta^\infty \eta^2(\beta_1)\, d\beta_1.$$

Hence

$$H[Q^\epsilon] - H[Q] = \epsilon\delta^\epsilon(0)\,; \quad P_\beta[Q^\epsilon] - P_\beta[Q] = -\epsilon\delta^\epsilon(\beta)/2,$$

$$\begin{aligned}
K[Q^\epsilon] - K[Q] &= -\int_0^\infty \delta^\epsilon(\beta) P_\beta d\beta + (\epsilon^2/4)\int_0^\infty \delta^\epsilon(\beta)\, d\beta \\
(8.01) \qquad &= \epsilon\delta^\epsilon(0) + \epsilon\int_0^\infty \frac{d}{d\beta}\,\delta^\epsilon P d\beta + (\epsilon^2/4)\int_0^\infty (\delta^\epsilon)^2 d\beta \\
&= \epsilon\delta^\epsilon(0) + \epsilon\Big[-2\int_0^\infty \eta\,|\,Q\,|\,P d\beta \\
&\quad - \epsilon\int_0^\infty \eta^2 P d\beta + (\epsilon/4)\int_0^\infty (\delta^\epsilon)^2 d\beta\Big].
\end{aligned}$$

For the latter equation the definition of $\delta^\epsilon(\beta)$ was used. Since $\eta \equiv 0$ for $\beta \geq \beta_2$ and $P > 0$ for $\beta \leq \beta_2$, we have $\int_0^\infty \eta\,|\,Q\,|\,P d\beta = \int_0^{\beta_2} \eta\,|\,Q\,|\,P d\beta > 0$.

It is clear, therefore, that $\epsilon$ can be made so small that the quantity in the brackets is negative. Hence

$$(8.02) \qquad K[Q^\epsilon]-K[Q]<\epsilon\delta^\epsilon(0).$$

Further we have, from the definition of $\eta(\beta)$,

$$\int_0^\infty [Q^\epsilon{}_\beta]^2 d\beta - \int_0^\infty Q_\beta{}^2 d\beta$$
$$=\epsilon[-2\int_{\beta_1}^{\beta_0}\eta_\beta\,|\,Q_\beta\,|\,d\beta+2\int_{\beta_0}^{\beta_2}\eta_\beta\,|\,Q_\beta\,|\,d\beta+\epsilon\int_{\beta_1}^{\beta_2}\eta_\beta{}^2 d\beta]$$
$$=\epsilon[-2\int_{\beta_1}^{\beta_2}|\,\eta_\beta\,|\,|\,Q_\beta\,|\,d\beta+\epsilon\int_{\beta_1}^{\beta_2}\eta_\beta{}^2 d\beta].$$

Evidently we can choose $\epsilon$ so small that also here the quantity in the brackets is negative; i. e.,

$$(8.03) \qquad \int_0^\infty [Q^\epsilon{}_\beta]^2 d\beta - \int_0^\infty Q_\beta{}^2 d\beta<0.$$

By combining (8.01), (8.02), and (8.03) we obtain

$$W[Q^\epsilon]-W[Q]<0$$

in contradiction with the minimum property of $Q$. This contradiction establishes Theorem 8.2.

Theorem 8.3 is equivalent to the statement: if $Q^0$ is any solution of $B^\kappa$ which does not vanish for $0\leq\beta<\infty$, then

$$W^\kappa[Q]\geq W^\kappa[Q^0]$$

for every admissible function $Q$ and the equality holds only for $Q=\pm Q^0$.

Let $P$ and $P^0$ be the functions corresponding to $Q$ and $Q^0$. We derive the identity

$$(8.04) \qquad \int_0^\infty (1-P^0)Q^2\rho^\kappa d\beta=2\int_0^\infty P_\beta{}^0 P_\beta d\beta.$$

The integral on the right-hand side exists in view of $\int_0^\infty [P_\beta{}^0]^2 d\beta<\infty$ and $\int_0^\infty P_\beta{}^2 d\beta<\infty$. To prove (8.04) we introduce the identity

$$\int_0^a (1-P^0)Q^2\rho^\kappa d\beta=2\int_0^a P_\beta{}^0 P_\beta d\beta+2(1-P^0)P_\beta|^a.$$

From Lemma 7.1 applied to $f=1-P^0$, $g=P$ it follows that the last term tends to zero when $a\to\infty$ on a special sequence and (8.04) is established. Formula (8.04) implies that $\int_0^\infty(1-P^0)Q^2\rho^\kappa d\beta$ exists for every admissible

$Q$; in view of $(1-P^0)\geq 0$, $|P^0|\leq 1+(1-P^0)$ and $\int_0^\infty Q^2\rho^\kappa d\beta<\infty$ we can infer

$$\int_0^\infty |P^0|\, Q^2\rho^\kappa d\beta < \infty. \tag{8.05}$$[34]

We now introduce the quadratic functional

$$T[Q]=\int_0^\infty Q_\beta{}^2 d\beta+\kappa(1+\nu)Q^2(0)-\int_0^\infty P^0Q^2\rho^\kappa d\beta, \tag{8.06}$$

which, in view of (8.05), is defined for admissible $Q$. From (8.04) we have

$$\begin{aligned} W[Q]&=T[Q]+\int_0^\infty P_\beta{}^2 d\beta-\int_0^\infty(1-P^0)Q^2\rho^\kappa d\beta \\ &=T[Q]+\int_0^\infty P_\beta{}^2 d\beta-2\int_a^\infty P^0{}_\beta P_\beta d\beta, \end{aligned} \tag{8.07}$$

and for $Q=Q^0$

$$W[Q^0]=T[Q^0]-\int_0^\infty (P^0{}_\beta)^2 d\beta.$$

Subtraction yields the identity

$$W[Q]-W[Q^0]=T[Q]-T[Q^0]+\int_0^\infty (P_\beta-P^0{}_\beta)^2 d\beta. \tag{8.08}$$

Theorem 8.3 is a consequence of (8.08) and

LEMMA 8.1. *For admissible $Q$*

$$T[Q]\geq 0, \tag{8.09}$$

*where the equality holds only for $Q=cQ^0$, $c=const.$*

Lemma 8.1 implies

$$T[Q^0]=0. \tag{8.09$_0$}$$

If Lemma 8.1 holds, (8.08) yields $W[Q]\geq W[Q^0]$, the equality holding only for $Q=cQ^0$, $P_\beta=P^0{}_\beta$ or $\int_0^\infty[Q^2-(Q^0)^2]\rho^\kappa d\beta=0$. Since $\rho^\kappa>0$, the last equality holds only when $Q=\pm Q^0$. Hence Theorem 8.3 is proved once the inequality (8.09) is established. This inequality states that $Q^0$ minimizes the quadratic functional $T[Q]$.

We proceed to establish Lemma 8.1. Our proof follows the line of

---

[34] From (8.05) and the admissibility conditions for $\kappa=0$, $\rho^0=1$, it follows that $\int_0^\infty[Q_\beta{}^2-P_\beta{}^2+PQ^2]d\beta$ is finite. For the solution $Q$, $P$ the integrand is constant according to (4.13). This constant is therefore zero, confirming the assumption made in (4.14).

Jacobi's treatment of the second variation; however, the infinite range of the independent variable and the singularity occurring in our problem require considerable modifications of the standard reasoning.[35] In view of $Q^0 > 0$ it is possible to introduce the function

$$\theta = Q/Q^0. \tag{8.10}$$

With this function $\theta$ the identity

$$T[Q] = \int_0^\infty \theta_\beta^2 (Q^0)^2 \, d\beta \tag{8.11}$$

holds, as we shall prove. The identity (8.11) implies that $T[Q] \geq 0$ and that the equality holds only if $\theta_\beta = 0$ (since $Q^0 \not\equiv 0$), that is, if $\theta \equiv c = \text{const.}$, or $Q = cQ^0$. Thus Lemma 8.1 follows from (8.11).

To prove (8.11) we apply Jacobi's identity (7.22) to $\omega = Q^0$ and obtain

$$\int_0^a [Q_\beta^2 + (Q^0)^{-1} Q^0_{\beta\beta} Q^2] d\beta = \int_0^a (Q^0)^2 \theta_\beta^2 d\beta + (Q^0)^{-1} Q^0_\beta Q^2 |_0^a,$$

or, in view of $Q^0_{\beta\beta} = -\rho^\kappa P^0 Q^0$ and $Q^0_\beta - \kappa(1+\nu) Q^0|^0 = 0$,

$$\int_0^a [Q_\beta^2 - P^0 Q^2 \rho^\kappa] d\beta + \kappa(1+\nu) Q^2(0) = \int_0^a (Q^0)^2 \theta_\beta^2 d\beta + (Q^0)^{-1} Q^0_\beta Q^2 |^a.$$

If it could be shown that the last term tends to zero as $a \to \infty$ on at least a special sequence, (8.11) would result in view of (8.06). It is doubtful whether such a direct attack is possible.

Instead, we proceed as follows: Since $\beta^{-1}[Q^0(\beta)]^2 \to 0$, according to Lemma 7.4, there is a special sequence $\beta \to \infty$ on which $(\beta^{-\frac{1}{2}} Q^0)_\beta \leq 0$, or $Q^0_\beta \leq \frac{1}{2}\beta^{-1} Q$, or $(Q^0)^{-1} Q^0_\beta Q^2 \leq \frac{1}{2}\beta^{-1} Q^2$. But $\beta^{-1} Q^2(\beta) \to 0$. If $a$ is taken on this sequence, the inequality

$$T[Q] \leq \int_0^\infty (Q^0)^2 \theta_\beta^2 d\beta \text{ results.} \tag{8.11$^-$}$$

Identity (8.11) will be established once the reverse inequality

$$T[Q] \geq \int_0^\infty (Q^0)^2 \theta_\beta^2 d\beta \text{ is proved.} \tag{8.11$^+$}$$

To this end we introduce the function $Q^\epsilon = Q^0 + \epsilon\beta$ and set

$$\theta^\epsilon = Q/Q^\epsilon \quad \text{for} \quad \epsilon > 0. \tag{8.10}$$

We then apply once more Jacobi's identity (7.22) with $\omega = Q^\epsilon$ and obtain

---

[35] **Our reasoning could be made the basis of a general theory extending Jacobi's treatment of the second variation to the case of singular or infinite end points.**

$$\int_0^a [Q_\beta^2 + (Q^\epsilon)^{-1}(Q^\epsilon{}_{\beta\beta})^2]d\beta = \int_0^a (Q^\epsilon)^2 (\theta^\epsilon{}_\beta)^2 d\beta + (Q^\epsilon)^{-1}(Q^\epsilon{}_\beta)^2 Q^2 |_0^a;$$

using the relation

$$-Q^\epsilon{}_{\beta\beta} = \rho^\kappa P^0 Q^0,\ (Q^\epsilon)^{-1}Q^0 = 1 - \epsilon\beta(Q^\epsilon)^{-1},\ \text{and}\ Q^\epsilon{}_\beta - (1+\nu)Q^\epsilon|^0 = \epsilon,$$

we have

$$\int_0^a [Q_\beta^2 - \rho^\kappa P^0 Q^2]d\beta + (1+\nu)Q^2(0)$$
$$= \int_0^a [(Q^\epsilon)^2(\theta^\epsilon{}_\beta)^2 - \epsilon\beta Q^{-1}\rho^\kappa P^0 Q^2]d\beta - \epsilon(Q^0)^{-1}Q^2|^0 + (Q^\epsilon)^{-1}Q^\epsilon{}_\beta Q^2|^a.$$

Since $\int_0^\infty Q_\beta{}^2 d\beta < \infty$, there is a special sequence $\beta \to \infty$ on which $\beta^{1/2}Q^0{}_\beta(\beta) \to 0$. On this sequence we have

$$\beta^{-1}Q^\epsilon = \beta^{-1}Q^0 + \epsilon \to \epsilon, \qquad Q^\epsilon{}_\beta = Q^0{}_\beta + \epsilon \to \epsilon, \qquad \beta^{-1}Q^2 \to 0$$

in view of Lemma 7.4; hence $(Q^\epsilon)^{-1}Q^\epsilon{}_\beta Q^2 \to 0$. If we let $a \to \infty$ on this special sequence we obtain the identity

$$(8.11)_\epsilon \quad T[Q] = \int_0^\infty [(Q^\epsilon)^2(\theta^\epsilon{}_\beta)^2 - \epsilon\beta(Q^\epsilon)^{-1}\rho^\kappa P^0 Q^2]d\beta - \epsilon(Q^0)^{-1}Q^2|^0.$$

Allowing $\epsilon$ to tend to zero in $(8.11)_\epsilon$ we have

$$\epsilon(Q^0)^{-1}Q^2|^0 \to 0, \qquad \int_0^a \epsilon\beta(Q^\epsilon)^{-1}\rho^\kappa P^0 Q^2 d\beta \to 0,$$

and

$$\left|\int_a^\infty \epsilon\beta(Q^\epsilon)^{-1}\rho^\kappa P^0 Q^2 d\beta\right| \leq \int_a^\infty \rho^\kappa \mid P^0 \mid Q^2 d\beta.$$

The last quantity can be made arbitrarily small by choice of $a$. Further we have

$$\int_0^b (Q^\epsilon)^2(\theta^\epsilon{}_\beta)^2 d\beta \to \int_0^b (Q^0)^2\theta_\beta{}^2 d\beta, \qquad \int_b^\infty (Q^\epsilon)^2(\theta^\epsilon{}_\beta)^2 d\beta \geq 0.$$

Hence $(8.11)_\epsilon$ yields in the limit

$$T[Q] \geq \int_0^b (Q^0)^2\theta_\beta{}^2 d\beta$$

and $(8.11)^+$ is established since $b$ is arbitrary. Hence (8.11) holds, Lemma 8.1 is proved and, with it, Theorem 8.3.

**9. Existence and asymptotic convergence.** In this section we prove the existence of the solutions of the minimum problem $M^\kappa$, including the asymptotic case $M^0$, and establish the convergence of the solutions for $\kappa > 0$ to the asymptotic solution ($\kappa = 0$) as $\kappa \to 0$. We apply direct methods similar

to those used for linear boundary value problems (cf. Courant-Hilbert [3], Volume II, Chapter VII).

We use the same formulation of the minimum problem $M^\kappa$ as in sec. 7. Functions $Q$ admissible with respect to the problem $M^\kappa$ in the sense of sec. 7 are here referred to as $\kappa$-admissible functions.

Our theorems are

THEOREM 9.1. *To every $\kappa \geq 0$ there exists at least one $\kappa$-admissible function $Q(\beta)$ for which $W^\kappa[Q]$ attains its minimum.*

Such a minimizing function will be denoted henceforth by $Q^\kappa(\beta)$; it is uniquely determined (Theorem 8.1) once the condition $Q^\kappa(0) \geq 0$ has been imposed.

THEOREM 9.2. *The minimizing functions $Q^\kappa(\beta)$ with $Q^\kappa(0) \geq 0$ tend as $\kappa \to 0$ to the minimizing function $Q^0(\beta)$ with $Q^0(0) \geq 0$ uniformly in each finite interval $0 \leq \beta \leq b < \infty$, and $W^\kappa[Q]$ tends to $W^0[Q^0]$.*

The proofs of the theorems are based upon a number of preliminary lemmas and inequalities.

We prove first

$$\left\{ \int_b^\infty Q^2 \rho^\kappa d\beta \right\}^2 \leq 4b^{-1} K^\kappa[Q] \tag{9.01}$$

for any $\kappa$-admissible $Q$ and $b > 0$. Due to $P_{\beta\beta} \geq 0$ (cf. (7.10)) the non-negative quantity $-P_\beta$ does not increase with increasing $\beta$. Hence

$$K^\kappa[Q] = \int_0^\infty P_\beta{}^2 d\beta \geq \int_0^b P_\beta{}^2 d\beta \geq b P_\beta{}^2(b) = (b/4) \left\{ \int_b^\infty Q^2 \rho^\kappa d\beta \right\}^2,$$

establishing (9.01).

Next we show that a constant $c > 0$ exists such that

$$H^\kappa[Q] \leq \tfrac{1}{2} D^\kappa[Q] + 2c\{K^\kappa[Q]\}^{1/2} \tag{9.02}$$

for any $\kappa$-admissible[36] $Q$ and all $\kappa \geq 0$.

To prove (9.02) we distinguish two cases: $\kappa \leq 1$, $\kappa \geq 1$.

*Case* 1 ($\kappa \leq 1$). We consider the successive inequalities

$$\begin{aligned} |Q(\beta_1) - Q(\beta_1 + \tfrac{1}{2})|^2 &\leq \left[ \int_{\beta_1}^{\beta_1 + 1/2} Q_\beta d\beta \right]^2 \leq \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta, \\ Q^2(\beta_1) &\leq 2Q^2(\beta_1 + \tfrac{1}{2}) + 2\,|Q(\beta_1) - Q(\beta_1 + \tfrac{1}{2})|^2 \\ &\leq 2Q^2(\beta_1 + \tfrac{1}{2}) + \int_0^\infty Q_\beta{}^2 d\beta. \end{aligned}$$

---

[36] The argument used to prove (9.02) could be so modified as to show that the admissibility condition on $H^\kappa$ is a consequence of the others and could therefore be omitted.

Integration with respect to $\beta_1$ yields

$$\int_0^{1/2} Q^2 d\beta \leq 2 \int_{1/2}^1 Q^2 d\beta + \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta.$$

Since $\rho^\kappa \leq 1$ and $\rho^\kappa = (1 + 2\kappa\beta)^{-6} \geq 3^{-6}$ for $\beta \leq 1$, $\kappa \leq 1$, we have

$$\int_0^{1/2} \rho^\kappa Q^2 d\beta \leq 2 \cdot 3^6 \int_{1/2}^1 \rho^\kappa Q^2 d\beta + \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta.$$

By adding $\int_{1/2}^\infty \rho^\kappa Q^2 d\beta$ to both sides we obtain

$$\int_0^\infty \rho^\kappa Q^2 d\beta \leq (1 + 2 \cdot 3^6) \int_{1/2}^\infty \rho^\kappa Q^2 d\beta + \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta.$$

We now use (9.01) with $b = \frac{1}{2}$, that is

$$\int_0^\infty \rho^\kappa Q^2 d\beta \leq 2\sqrt{2} \left\{ \int_0^\infty P_\beta{}^2 d\beta \right\}^{1/2}$$

and thus obtain

$$\int_0^\infty \rho^\kappa Q^2 d\beta \leq 2c \left\{ \int_0^\infty P_\beta{}^2 d\beta \right\} + \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta$$

where $c = \sqrt{2}(1 + 2 \cdot 3^6)$. This establishes (9.02) for $\kappa \leq 1$.

*Case* 2 ($\kappa \geq 1$). We set

$$\omega = (1 + \beta)^{-1}(1 + 2\beta)$$

in Jacobi's identity (7.22) noting that

$$\omega_\beta = (1 + \beta)^{-2}, \quad \omega^{-1}\omega_\beta = (1 + \beta)^{-1}(1 + 2\beta)^{-1}, \quad \omega_{\beta\beta} = -2(1 + \beta)^{-3},$$

and

$$\omega^{-1}\omega_{\beta\beta} = 2(1 + \beta)^{-2}(1 + 2\beta)^{-1} \geq 2(1 + 2\kappa\beta)^{-6} = 2\rho^\kappa(\beta).$$

The result is

$$\int_0^\infty (Q^2 - 2\rho^\kappa Q^2) d\beta \geq \int_0^\infty \omega^2(\omega^{-1}Q)_\beta{}^2 d\beta + \omega^{-1}\omega_\beta Q^2 \Big|_0^\infty.$$

Since

$$\omega^{-1}\omega_\beta \big|^0 = 1, \quad \beta\omega^{-1}\omega_\beta \big|^\infty = 0, \quad \text{and} \quad \beta^{-1}Q^2 \big|^\infty = 0,$$

according to Lemma 7.4, the right member of the inequality is non-negative, i. e.,

$$\int_0^\infty \rho^\kappa Q^2 d\beta \leq \tfrac{1}{2} \int_0^\infty Q_\beta{}^2 d\beta.$$

This establishes (9.02) for $\kappa \geq 1$ and hence generally.

From (9.02) we deduce the inequality

$$W^\kappa[Q] \geq D^\kappa[Q] + \{(K^\kappa[Q])^{1/2} - c\}^2 - c^2, \tag{9.03}$$

which implies

LEMMA 9.1. *For $\kappa$-admissible functions $Q$, $W^\kappa[Q]$ has a lower bound $(-c^2)$ independent of $\kappa$.*

From (9.03) we obtain

$$
\begin{aligned}
&D^\kappa[Q] \leq W^\kappa[Q] + c^2, \\
(9.04)\qquad &K^\kappa[Q] \leq \{c + (W^\kappa[Q] + c^2)^{1/2}\}^2, \text{ and from (9.02)} \\
&H^\kappa[Q] \leq 3c^2 + W^\kappa[Q] + 2c(W^\kappa[Q] + c^2)^{1/2}.
\end{aligned}
$$

Consider a set of $\kappa$-admissible functions $Q$ (with $\kappa$ not necessarily fixed) for which $W^\kappa$ has an upper bound $M$. We conclude from (9.04)

LEMMA 9.2. *An upper bound $M$ for $W^\kappa$ implies upper bounds for $D^\kappa$, $H^\kappa$, and $K^\kappa$ which depend upon $M$ but not on $\kappa$.*

In what follows we shall make frequent use of the following well-known lemmas:

LEMMA A. *If $f_s(\beta)$ is a sequence of non-negative continuous functions, defined for $0 \leq \beta < \infty$, which converge to a limit function $f_0(\beta)$ uniformly in every finite interval, then $f_0(\beta)$ is continuous, non-negative, and*

$$\int_0^\infty f_0(\beta)\,d\beta \leq \underline{\lim} \int_0^\infty f_s(\beta)\,d\beta.$$

LEMMA B. *If, in addition, to every $\epsilon > 0$ a quantity $b = b(\epsilon)$ exists such that*

$$\int_b^\infty f_s(\beta)\,d\beta \leq \epsilon \textit{ for all } s, \textit{ then}$$

$$\int_0^\infty f_s(\beta)\,d\beta \to \int_0^\infty f_0(\beta)\,d\beta.$$

We proceed to state the following basic

LEMMA 9.3. *Let $\kappa_m$ be a sequence of values of $\kappa$ tending to a finite value $\kappa_*$, and let $Q_m$ be a sequence of $\kappa_m$-admissible functions for which $W^{\kappa_m}[Q_m]$ is bounded. Then there exists a subsequence $Q_s$ converging uniformly in every finite interval to a $\kappa_*$-admissible function $Q$ such that*

$$\underline{\lim}\, D^{\kappa_s}[Q_s] \geq D^{\kappa_*}[Q_*], \tag{9.05}$$

$$\underline{\lim}\, H^{\kappa_s}[Q_s] = H^{\kappa_*}[Q_*], \tag{9.06}$$

$$\underline{\lim}\, K^{\kappa_s}[Q_s] \geq K^{\kappa_*}[Q_*], \tag{9.07}$$

$$\underline{\lim}\, W^{\kappa_s}[Q_s] \geq W^{\kappa_*}[Q_*]. \tag{9.08}$$

*Proof.* Lemma 9.2 shows that $D^{\kappa_m}[Q_m]$, $H^{\kappa_m}[Q_m]$, and $K^{\kappa_m}[Q_m]$ are bounded. A bound for $D^{\kappa_m}[Q_m]$ (cf. (7.08)) implies a bound for the sequence $\int_0^\infty Q^2{}_{m\beta}(\beta)\,d\beta$. Hence, as is well known, there exists a subsequence

converging uniformly in every finite interval $0 \leq \beta \leq b$ to a continuous limit function $Q_*$ with $L^2$-integrable derivative and such that

$$\underline{\lim} \int_0^\infty [Q_{s\beta}]^2(\beta)\, d\beta \geq \int_0^\infty [Q_{*\beta}]^2 d\beta.$$

This inequality yields (9.05), in view of (7.08) and $Q_s(0) \to Q_*(0)$. In order to establish (9.06) we use (9.01) in the form

$$\int_0^\infty \rho^\kappa Q_s^2 d\beta \leq 2b^{-1/2} K^{\kappa_s}[Q_s].$$

Since $K^{\kappa_s}$ is bounded, the right member can be made arbitrarily small by proper choice of $b$. Hence we may apply Lemma B with $f_s = \rho^{\kappa_s} Q_s^2$ and thus obtain (9.06). By virtue of the definition (7.10) of $P_\beta$ we have

$$| P_\beta^{\kappa_s}[Q_s] - P_\beta^*[Q_*] | \leq 2 \, | H^{\kappa_s}[Q_s] - H^{\kappa_*}[Q_*] |$$

and (9.06) yields the uniform convergence of $P_\beta^{\kappa_s}[Q_s]$ to $P_\beta^{\kappa_*}[Q_*]$. Hence we may apply Lemma A to $K^{\kappa_s}$; the result is (9.07).

As a consequence of (9.05), (9.06), and (9.07) $Q_*$ is $\kappa_*$-admissible and, in addition, (9.08) holds.

We are now in a position to prove Theorem 9.1. We turn, therefore, to the problem of minimizing $W^\kappa[Q]$ by a $\kappa$-admissible function $Q$. From Lemma 9.1 we know that the g. l. b. $\omega^\kappa$ of $W^\kappa[Q]$ is finite; hence there exists a minimizing sequence, i. e., a sequence of $\kappa$-admissible functions $Q_m$ for which $W^\kappa[Q_m]$ has as limit $\omega^\kappa$. We now apply Lemma 9.3 with $\kappa_m = \kappa_* = \kappa$; it yields the existence of a subsequence $Q_s$ and a $\kappa$-admissible function[37] $Q = Q_*$ for which (cf. (9.08))

$$W^\kappa[Q] \leq \underline{\lim}\, W^\kappa[Q_s].$$

Since the right member here is the g. l. b. $\omega^\kappa$ of $W^\kappa$, the equality must hold. Hence $Q$ solves the minimum problem $M^\kappa$. This proves Theorem 9.1.

We add the remark that the minima $\omega^\kappa$ of $W^\kappa$ have the common *upper* bound zero, i. e.,

$$\omega^\kappa \leq 0, \tag{9.09}$$

which follows immediately from $W^\kappa[0] = 0$ since $Q \equiv 0$ is an admissible function.

Before proving Theorem 9.2 we establish two lemmas:

---

[37] If we had required the existence of continuous second derivatives for admissibility it would have been necessary at this point to prove that the limit function $Q$ has this property—otherwise $Q$ could not be identified as the solution of $M^\kappa$. Our procedure, which requires only $L^2$-integrable derivatives for admissibility, makes it possible to separate the problems of existence and continuous differentiability of the solution of $M^\kappa$.

LEMMA 9.4. *A 0-admissible function $Q(\beta)$ is also $\kappa$-admissible for every $\kappa > 0$.*

This is an immediate consequence of $\rho^\kappa \leq \rho^0 = 1$ and

$$P_\beta{}^\kappa[Q] \leq P_\beta{}^0[Q], \tag{9.10}$$

in view of definitions (7.07) to (7.10).

LEMMA 9.5. *For any 0-admissible function $Q(\beta)$:*

$$W^\kappa[Q] \to W^0[Q] \text{ as } \kappa \to 0. \tag{9.11}$$

*Proof.* From (7.08) it is obvious that

$$D^\kappa[Q] \to D^0[Q] \text{ as } \kappa \to 0. \tag{9.12}$$

We apply Lemma B to

$$H^0[Q] - H^\kappa[Q] = \int_0^\infty Q^2(\beta)(1-\rho^\kappa)\,d\beta,$$

observing that

$$\int_b^\infty Q^2(\beta)(1-\rho^\kappa)\,d\beta \leq \int_b^\infty Q^2(\beta)\,d\beta$$

can be made arbitrarily small by proper choice of $b$. It follows that

$$H^\kappa[Q] \to H^0[Q] \text{ as } \kappa \to 0. \tag{9.13}$$

The latter implies, in view of (7.10), that

$$P_\beta{}^\kappa[Q] \to P_\beta{}^0[Q] \text{ as } \kappa \to 0. \tag{9.14}$$

We now consider

$$K^0[Q] - K^\kappa[Q] = \int_0^\infty \{(P_\beta{}^0)^2 - (P_\beta{}^\kappa)^2\}\,d\beta. \tag{9.15}$$

The right member, in view of (9.10), has a positive integrand and

$$\int_b^\infty \{(P_\beta{}^0)^2 - (P_\beta{}^\kappa)^2\}\,d\beta < \int_b^\infty (P_\beta{}^0)^2\,d\beta$$

can be made small. This fact and (9.14) permit us to apply Lemma B to (9.15) with the result

$$K^\kappa[Q] \to K^0[Q] \text{ as } \kappa \to 0. \tag{9.16}$$

Relations (9.12), (9.13), and (9.16) lead immediately to (9.11).

We now take *any* sequence of positive values $\kappa$ tending to zero and solutions $Q^\kappa$ of the corresponding minimum problems $M^\kappa$ (which exist according to Theorem 9.1). We may assume $Q^\kappa(0) \geq 0$. As remarked above (cf. (9.09)), the values $W^\kappa[Q^\kappa] = \omega^\kappa$ have an upper bound. We can therefore apply Lemma 9.3 to the sequence $Q^\kappa$ with $\kappa_* = 0$. The lemma insures the existence of a subsequence converging in the sense of the lemma to a 0-admissible limit

function $Q_0$. From now on $Q^\kappa$ refers to such a subsequence. From (9.08) we have, in particular,

$$\underline{\lim}\, W^\kappa[Q^\kappa] \geq W^0[Q_0]. \tag{9.17}$$

We proceed to show that $Q_0$ *solves* the minimum problem $M^0$. This minimum problem, according to Theorem 9.1, has a solution $Q^0$ for which (Lemma 9.5)

$$\lim W^\kappa[Q^0] = W^0[Q^0]. \tag{9.18}$$

As a consequence of the minimum properties of $Q^0$ and $Q^\kappa$ we have

$$\omega^0 = W^0[Q^0] \leq W^0[Q_0] \quad \text{and} \tag{9.19}$$

$$\omega^\kappa = W^\kappa[Q^\kappa] \leq W^\kappa[Q^0]. \tag{9.20}$$

Here we have made use of the fact that $Q^0$ is $\kappa$-admissible (Lemma 9.4). From (9.20) we find

$$\overline{\lim}\, W^\kappa[Q^\kappa] \leq \overline{\lim}\, W^\kappa[Q^0]. \tag{9.21}$$

Successive consideration of (9.21), (9.18), (9.19), and (9.17) yields

$$\overline{\lim}\, W^\kappa[Q^\kappa] \leq W^0[Q^0] \leq W^0[Q_0] \leq \underline{\lim}\, W^\kappa[Q^\kappa] \tag{9.22}$$

and this obviously implies the equality

$$W^0[Q^0] = W^0[Q_0]. \tag{9.23}$$

Since $W^0[Q^0]$ is the g. l. b. of $W^0$ it follows that the function $Q = Q_0$ is a solution of the minimum problem $M^\kappa$ for $\kappa = 0$ with $Q(0) \geq 0$.

From Theorem 8.1 we know that the minimum problem $M^0$ has at most one solution $Q$ with $Q(0) \geq 0$. Therefore $Q^0 \equiv Q_0$, i. e., all convergent sequences $Q^\kappa$ converge to the same limit function. If a sequence has the property that every subsequence contains a convergent subsequence with limit $L$ and if $L$ is the same for all such convergent subsequences, then the original sequence itself converges to $L$. Therefore we conclude in our case that the solutions $Q^\kappa(\beta)$ with $Q^\kappa(0) \geq 0$ of the minimum problems $M^\kappa$ converge, as $\kappa \to 0$, to the unique solution $Q^0(\beta)$ with $Q^0(0) \geq 0$ of the minimum problem $M^0$. This completes the proof of Theorem 9.2.

**10. Limit state in the interior of plate.** While the limit procedure of sec. 9 concerns the *boundary layer,* we deal in this section with the limit procedure in the *interior* of the plate as $\lambda \to \infty$. These two limit procedures have already been discussed and contrasted with each other in sec. 4.

We summarize the discussion of sec. 4 concerning the interior limit process, but we use the notation of sec. 5. By multiplying (5.10) by $\kappa^3$ and (5.12), (5.13) by $\kappa^2$ $(\kappa = \lambda^{-1})$ we obtain

$$\kappa^2 G(\kappa q) + (\kappa^2 p)(\kappa q) = 0, \tag{10.01}$$

(10.02) $$G(\kappa^2 p) = (\kappa q)^2/2, \quad \text{and}$$

(10.03) $$\kappa^2 p = 1 \quad \text{for} \quad r = 1.$$

When $\kappa \to 0$, the following equations (cf. 4.01) result

(10.04) $$\hat{p}\hat{q} = 0,$$

(10.05) $$G(\hat{p}) = \hat{q}^2/2, \quad \text{and}$$

(10.06) $$\hat{p} = 1 \quad \text{for} \quad r = 1, \quad \text{where}$$

(10.07) $$\hat{p} = \lim (\kappa^2 p), \quad \hat{q} = \lim (\kappa q) \quad \text{for} \quad \kappa \to 0.$$

One expects that the limit equations (10.04) and (10.05) will be satisfied. The only admissible solution of them is $\hat{p} = \text{const.}$, $\hat{q} = 0$. It was pointed out in sec. 4 that the proper constant value for $\hat{p}$ is not to be taken from (10.06) but rather from the asymptotic solution of the boundary layer problem, i.e., $\hat{p} = P(\infty)$. Since $P(\infty)$ is negative, this means that the radial membrane stress tends to a negative constant in every interior region.

Precisely, we prove the following

THEOREM 10.1. *As $\kappa \to 0$ the limit relations*

(10.08) $$\kappa^2 p_\lambda(r) \to P^0(\infty),$$ (10.09) $$\kappa q_\lambda(r) \to 0$$

*hold uniformly in every interval $0 \leq r \leq r_0 < 1$. Here $q_\lambda(r)$ is the solution of the minimum problem $M_\lambda$ (cf. sec. 5) and $p_\lambda$ is given by* (5.07) *and* (5.06).

Once Theorem 10.1 has been established, it is clear that the limit procedure discussed above leads to correct results for those solutions of $B_\lambda$ which do not vanish (cf. Theorem 8.3).

With the notation of sec. 7 the convergence relations (10.08) and (10.09) become

(10.10) $$P^\kappa(\kappa^{-1}\zeta) \to P^0(\infty),$$ (10.11) $$\kappa^{-1} Q^\kappa(\kappa^{-1}\zeta) \to 0;$$

where

(10.12) $$\zeta = \tfrac{1}{2}(r^{-2} - 1) = \kappa\beta.$$

Theorem 10.1 is evidently proved once these relations are shown to hold uniformly in every interval $0 < \zeta_0 \leq \zeta$. $Q^\kappa(\beta)$ is the solution of the minimum problem $M^\kappa$. In what follows the superscript $\kappa$ on the quantities $P$ and $Q$ is omitted where no confusion could result.

We make use of various properties of the functions $Q$ and $P$. From Theorem 8.2 we have

(10.13) $$Q^\kappa(\beta) \geq 0.$$

We show next that there exists a value $\beta_0$ of $\beta$ such that

$$P^0(\beta_0) < 0. \tag{10.14}$$

This can be seen indirectly as follows. If $P^0(\beta) \geq 0$ for all $\beta$, (10.13) and (7.12) for $\kappa = 0$ show that $Q^0{}_{\beta\beta}$ would be $\leq 0$. Consequently the function $Q^0(\beta) \not\equiv 0$ would be non-decreasing or would become negative infinite. But both cases are impossible in view of the admissibility condition

$$\int_0^\infty (Q^0)^2 d\beta < \infty.$$

Since $P^\kappa(\beta_0)$ and $Q^\kappa(\beta_0)$ converge to $P^0(\beta_0)$ and $Q^0(\beta_0)$ as $\kappa \to 0$ (Theorem 9.2) there exist positive values $\kappa_0$, $c$, $c_1$, and $d$ such that for $\kappa \leq \kappa_0$

$$0 < c \leq -P^\kappa(\beta_0) \leq c_1, \quad \text{and} \tag{10.15}$$

$$Q^\kappa(\beta_0) \leq d \tag{10.16}$$

hold. From now on we assume always $\kappa \leq \kappa_0$ and $\beta \geq \beta_0$.

Our subsequent discussion is based on the following formulas, which hold for all $\kappa \geq 0$:

$$-P_\beta = \tfrac{1}{2} \int_\beta^\infty \bar{Q}^2 \bar{\rho}^\kappa d\bar{\beta}, \tag{10.17}$$

$$P(\beta_1) - P(\beta_2) = \tfrac{1}{2} \int_{\beta_1}^{\beta_2} \bar{Q}^2 (\bar{\beta} - \beta_1) \bar{\rho}^\kappa d\bar{\beta} - (\beta_2 - \beta_1) P_\beta(\beta_2), \tag{10.18}$$

$$-Q_\beta(\beta) = \int_\beta^\infty (-\bar{P}) \bar{Q} \bar{\rho}^\kappa d\bar{\beta}, \quad \text{and} \tag{10.19}$$

$$\begin{aligned} Q(\beta_1) - Q(\beta_2) \\ = \int_{\beta_1}^{\beta_2} (-\bar{P}) \bar{Q} (\bar{\beta} - \beta_1) \bar{\rho}^\kappa d\bar{\beta} - (\beta_2 - \beta_1) Q_\beta(\beta_2), \end{aligned} \tag{10.20}$$

where the bar indicates that $\beta$ has been replaced by $\bar{\beta}$.

Formula (10.17) is the definition of $P_\beta$ (sec. (7.10)). Formula (10.18) is obtained by integrating (10.17) from $\beta_1$ to $\beta_2$ and applying integration by parts to the right-hand side. Formula (10.19) results from the differential equation (7.12) by integration from $\beta$ to $\infty$ and use of the fact that $Q_\beta(\beta) \to 0$ if $\beta \to \infty$ at least on a special sequence, the latter following from the admissibility condition $D^\kappa[Q] < \infty$ (cf. (7.08)). Formula (10.20) follows from (10.19) in the same way as (10.18) was obtained from (10.17).

From (10.17) we have

$$-P_\beta \geq 0; \text{ from (10.18) and (10.21)} \tag{10.21}$$

$$-P(\beta_1) \leq -P(\beta_2) \quad \text{for} \quad \beta_1 \leq \beta_2, \tag{10.22}$$

and, in addition, in view of (10.15),

$$-P(\beta) \geq c > 0, \qquad \beta \geq \beta_0. \tag{10.23}$$

Correspondingly we obtain from (10.19), (10.20), and (10.16), in view of (10.23) and (10.13):

(10.24) $$Q_\beta \leq 0,$$

(10.25) $$Q(\beta_1) \geq Q(\beta_2) \quad \text{for} \quad \beta_1 \leq \beta_2, \quad \text{and}$$

(10.26) $$Q(\beta) \leq d, \qquad \beta \geq \beta_0.$$

In view of (10.23), (10.25), and (10.24) we decrease the right member of (10.20) when we replace $-\bar{P}$ by $c$, $\bar{Q}$ by $Q(\beta_2)$ and omit the second term. We have thus

$$Q(\beta_1) - Q(\beta_2) \geq cQ(\beta_2) \int_{\beta_1}^{\beta_2} (\bar{\beta} - \beta_1)\bar{\rho}^{\kappa} d\bar{\beta}.$$

The left member of this inequality, in view of (10.26) and (10.13), is $\leq d$. Hence we obtain

(10.27) $$Q(\beta_2) \leq dc^{-1}\left[\int_{\beta_1}^{\beta_2} (\bar{\beta} - \beta_1)\bar{\rho}^{\kappa} d\bar{\beta}\right]^{-1}.$$

For $\kappa = 0$, $(\rho^0 = 1)$, (10.27) yields, with $\beta_1 = \beta_0$, $\beta_2 = \beta$,

(10.28) $$Q^0(\beta) \leq 2dc^{-1}(\beta - \beta_0)^{-2}.$$

For $\kappa > 0$ we estimate the bracket on the right side of (10.27). We introduce the function $\sigma(\beta_1, \beta_2)$ by the formula

(10.29) $$\sigma(\beta_1, \beta_2) = \int_{\beta_1}^{\beta_2} (\bar{\beta} - \beta_1)\bar{\rho}^1 d\bar{\beta},$$

from which, in view of $\rho^\kappa(\beta) = \rho^1(\kappa\beta)$ (cf. (7.06)), we have

(10.30) $$\int_{\beta_1}^{\beta_2} (\bar{\beta} - \beta_1)\bar{\rho}^{\kappa} d\bar{\beta} = \kappa^{-2}\sigma(\kappa\beta_1, \kappa\beta_2).$$

We insert (10.30) in (10.27) with $\beta_1 = \beta_0$, replace $\beta_2$ by $\kappa^{-1}\zeta$ (cf. (10.12)), and obtain for $\kappa \leq \zeta/\beta_0$

(10.31) $$\kappa^{-2}Q(\kappa^{-1}\zeta) \leq dc^{-1}/\sigma(\kappa\beta_0, \zeta).$$

We have already restricted $\zeta$ to the interval $0 < \zeta_0 \leq \zeta$, and, if we restrict $\kappa$ by $\kappa \leq \zeta_0/2\beta$ (as we may do), we have $\sigma(\kappa\beta_0, \zeta) \geq \sigma(\zeta_0/2, \zeta_0)$. Hence we derive from (10.31) the inequality

(10.32) $$\kappa^{-2}Q(\kappa^{-1}\zeta) \leq dc^{-1}/\sigma(\zeta_0/2, \zeta_0).$$

The left member is bounded as $\kappa$ tends to zero for every $\zeta_0 > 0$. This obviously proves the convergence relation (10.11).

We turn now to the proof of the relation (10.10). An immediate consequence of (10.26) and (10.23) is

$$Q^2 \leq dc^{-1}(-PQ);$$

insertion of this in (10.18) after replacing $P_\beta$ in accordance with (10.17), and comparison with (10.20) and (10.19) yields for $\beta_2 > \beta_1$

$$P(\beta_1) - P(\beta_2) \leq \tfrac{1}{2}dc^{-1}[Q(\beta_1) - Q(\beta_2)]. \tag{10.33}$$

The right member is $\leq \frac{1}{2}dc^{-1}Q(\beta_1)$ in view of (10.25). Hence, when $\beta_2 \to \infty$, $-P(\beta_2)$ tends to a finite limit $-P(\infty)$ since it increases monotonically (cf. (10.22)).

We consider the inequality

$$\begin{aligned} |P^\kappa(\kappa^{-1}\zeta) - P^0(\infty)| \leq |P^\kappa(\kappa^{-1}\zeta) - P^\kappa(\beta_1)| + |P^\kappa(\beta_1) - P^0(\beta_1)| \\ + |P^0(\beta_1) - P^0(\infty)|. \end{aligned}$$

The first term on the right is to be estimated by (10.33) and

$$|Q^\kappa(\kappa^{-1}\zeta) - Q^\kappa(\beta_1)| \leq |Q^\kappa(\kappa^{-1}\zeta)| + |Q^\kappa(\beta_1) - Q^0(\beta_1)| + |Q^0(\beta_1)|.$$

The resulting inequality is

$$\begin{aligned} |P^\kappa(\kappa^{-1}\zeta) - P^0(\infty)| &\leq \tfrac{1}{2}dc^{-1}|Q^\kappa(\kappa^{-1}\zeta)| \\ &+ \tfrac{1}{2}dc^{-1}|Q^\kappa(\beta_1) - Q^0(\beta_1)| + \tfrac{1}{2}dc^{-1}|Q^0(\beta_1)| + |P^\kappa(\beta_1) - P^0(\beta_1)| \\ &+ |P^0(\beta_1) - P^0(\infty)|. \end{aligned} \tag{10.34}$$

Since $P^0(\beta_1) \to P^0(\infty)$ and $Q^0(\beta_1) \to 0$ as $\beta_1 \to \infty$ (cf. (10.28)) we can choose $\beta_1$ so large that the third and fifth terms on the right side of (10.34) are arbitrarily small. Since $P^\kappa(\beta_1) \to P^0(\beta_1)$ and $Q^\kappa(\beta_1) \to Q^0(\beta_1)$ (Theorem 9.2), and since $Q^\kappa(\kappa^{-1}\zeta) \to 0$ (cf. (10.32)) as $\kappa \to 0$, we can choose $\kappa$ so small that the remaining terms on the right side of (10.34) become arbitrarily small. This establishes the convergence relation (10.10) and the proof of Theorem 10.1 is completed.

## APPENDIX.

**Perturbation method and E. Schmidt's bifurcation theory.** The perturbation method of sec. 2 finds its theoretical justification in its close relation to the bifurcation theory developed by E. Schmidt for a class of non-linear integral equations (cf. [14] and the literature given there). In this appendix we shall reduce our problem to an integral equation of a similar type. A certain singularity occurring in our problem does not interfere with the applicability of E. Schmidt's procedure, of which we give a brief indication.

We begin by considering the linear boundary value problem

$$(\mathrm{B})_f \qquad (G + \lambda_0^2)q = -f, \quad rq_r + (1+\nu)q|^1 = 0, \quad qDq < \infty, \quad qHq < \infty,$$

where the operator $G$ and the forms $D$ and $H$ are defined in sec. 5 and $f = f(r)$ is any given function. The corresponding homogeneous problem

$(f \equiv 0)$ is denoted by $(\mathrm{B})_0$; we assume $\lambda_0{}^2$ to be the lowest eigenvalue of $(\mathrm{B})_0$ (cf. sec. 5).

Our non-linear boundary value problem can also be written in the form $(\mathrm{B})_f$ by taking for $f$ the functional

$$f = \mu q - O(q^2)q, \tag{I. 01}$$

where the functional $O$ is defined by

$$O(q^2) = \tfrac{1}{2}\int_r^1 r_1{}^{-3}\int_0^{r_1} q^2(r_2)r_2{}^3 dr_2 dr, \tag{I. 02}$$

while $\mu$ is given by

$$\mu = \bar{p} - \lambda_0{}^2 = \bar{p} - p^0 \tag{I. 03}$$

and represents the excess of the applied load $\bar{p}$ over the critical load $p^0$. The problem thus defined will be referred to as (B) from now on. That (B) is our original problem can be seen as follows: instead of (5. 07) we have $\bar{p} - p = O(q^2)$, hence (I. 01), in view of (I. 03), is the same as $f = (p - \lambda_0{}^2)q$ and the differential equation in (B) becomes $Gq + pq = 0$, i. e., equation (5. 10).

We proceed to construct the solutions of the linear problem $(\mathrm{B})_f$. If $\lambda_0{}^2$ were not an eigenvalue of $(\mathrm{B})_0$, the problem $(\mathrm{B})_f$ could be solved by means of the Green's function; in our case, however, $\lambda_0{}^2$ being an eigenvalue, an improper Green's function must be used to represent the solutions. Let $j(r) = ar^{-1}J_1(\lambda_0 r)$ (cf. (2. 3)) be the first eigenfunction of $(\mathrm{B})_0$ where the constant $a$ is determined by

$$jHj = 1. \tag{I. 04}$$

The improper Green's function $s(r,\rho)$ (cf. Courant-Hilbert [3] Bd. I, Kap. V, § 14, 2) is then characterized as the solution of the boundary value problem $(\mathrm{B})_f$ with $f = -j(r)j(\rho)$, which is continuous but has for $r = \rho$ the jump singularity

$$\left[\frac{d}{dr}s(r,\rho)\right]_{\rho-}^{\rho+} = -1.$$

We note that $s(r,\rho)$ is symmetric in $r$ and $\rho$:

$$s(r,\rho) = s(\rho,r).$$

Such a function can be determined explicitly:

$$\begin{aligned} s(r,\rho) = -(\pi/2a^2)n(r)j(\rho) &+ \tfrac{1}{2}[m(r)j(\rho) + j(r)m(\rho)] \\ &+ bj(r)j(\rho) \end{aligned}$$

for $r \geq \rho$, and, by symmetry, for $r \leq \rho$. Here $j(r)$ is the function defined above, and $n(r) = ar^{-1}Y_1(\lambda_0 r)$, $m(r) = a\lambda_0{}^{-1}J_2(\lambda_0 r)$, where $Y_1$ is the Bessel

function of first order and second kind and $J_2$ that of second order and first kind. The constant $b$ is arbitrary and indicates that $s(r,\rho)$ is determined only within a multiple of $j(r)j(\rho)$.

With $s(r,\rho)$ we construct the operation

$$Sh = \int_0^1 s(r,\rho)h(\rho)\rho^3 d\rho$$

for any function $h(r)$ with $hHh < \infty$. The function $Sh$ satisfies the differential equation

$$(G + \lambda^2)Sh = -h + j(jHh), \tag{I. 05}$$

as can be inferred from the above properties of the Green's function $s(r,\rho)$; in addition, $q = Sh$ satisfies the other conditions of problem $(\mathrm{B})_f$. For a function $f$ orthogonal to $j$, i. e., satisfying

$$jHf = 0, \tag{I. 06}$$

the function $q = Sf$ is then a solution of $(\mathrm{B})_f$. Since such a solution of $(\mathrm{B})_f$ is unique within a multiple of $j$, we obtain *all* solutions of $(\mathrm{B})_f$ in the form

$$q = \epsilon j + Sf \quad \text{for constant } \epsilon. \tag{I. 07}$$

Upon applying the operation $S$ to $h = j$ we obtain $(G + \lambda_0^2)Sj = 0$, indicating that $Sj$ is an eigenfunction of $(\mathrm{B})_0$ and hence that $Sj$ is a multiple of $j$, say $Sj = cj$. After replacing $s(r,\rho)$ by $s(r,\rho) - cj(r)j(\rho)$ (which amounts to fixing the constant $b$ in the explicit expression for $s(r,\rho)$), the relation

$$Sj = 0, \tag{I. 08}$$

follows from the definition of the operation $S$ and (I. 04). We assume (I. 08) to hold from now on. An immediate consequence of (I. 08) and the symmetry of $s(r,\rho)$ is

$$jHSh = 0, \tag{I. 09}$$

indicating that the function $Sh$ is orthogonal to $j$ for arbitrary $h$.

We now turn to problem (B) where $f$ is given as a functional in $q$ through (I. 01). Upon inserting $f$ as given by (I. 01) into (I. 07) we obtain

$$q = \epsilon j + \mu Sq - S\{qO(q^2)\}, \tag{I. 10}$$

an integral equation which must be satisfied by every solution $q$ of (B). However, a function satisfying (I. 10) is a solution of (B) only if the condition (I. 06) is satisfied, which, upon using (I. 01), takes the form

$$jHq - jH\{qO(q^2)\} = 0.$$

This "bifurcation equation" is to be considered as a relation between $\mu$ and $\epsilon$. Bv virtue of (I. 07), (I. 04), and (I. 09) it reduces to

(I. 11) $$\mu\epsilon - jH\{qO(q^2)\} = 0.$$

As we are interested in solutions in the neighborhood of $\epsilon = 0$, $q \equiv 0$, it is appropriate to introduce a new function $t$ through

(I. 12) $$q = \epsilon j + \epsilon t, \text{ where } t, \text{ of course, depends on } \epsilon.$$

In view of (I. 06) equation (I. 10) becomes

(I. 13) $$t = \mu St - \epsilon^2 S\{(j+t)O(j+t)^2\},$$

while the bifurcation equation (I. 11) assumes the form

(I. 14) $$\epsilon[\mu - \epsilon^2 jH\{(j+t)O(j+t)^2\}] = 0.$$

We exclude the trivial solution $\epsilon = 0$, $t \equiv 0$ valid for any $\mu$. Then (I. 14) reduces to

(I. 15) $$\mu = \epsilon^2 jH\{(j+t)O(j+t)^2\}.$$

Once the relations (I. 13) and (I. 15) have been obtained, the procedure of E. Schmidt can be applied: 1) $\epsilon$ and $\mu$ are considered as independent parameters and the integral equation (I. 13) is solved[38] on this basis; 2) the solution $t(r; \epsilon^2, \mu)$ of (I. 13) is inserted in the relation (I. 15); $\mu$ is then considered a function $\mu(\epsilon^2)$ of $\epsilon^2$ through the resulting transcendental equation, the bifurcation equation; 3) $t(r; \epsilon^2, \mu(\epsilon^2))$ inserted in (I. 12) determines finally the solution of problem (B).

Following the procedure of E. Schmidt, it is possible to show that (I. 13) has a solution which is a *power* series in $\mu$ and $\epsilon^2$ for small enough $\mu$ and $\epsilon$. The relation between $\mu$ and $\epsilon$ derived from (I. 15) will then be such that $\mu$ can be expressed as a power series in $\epsilon^2$, starting with the term $\epsilon^2 jHjO(j^2)$ which does not vanish.[39] That is: $\mu = \bar{p} - p^0$ is a power series in $\epsilon^2$ and $q$ a series in odd powers of $\epsilon$. This was *assumed* in working with the perturbation method (sec. 2).

NEW YORK UNIVERSITY.

---

[38] Note that $O(q^2)$ and $Sq$ are continuous even at $r = 0$ if $q$ has this property. Hence iterations on (I. 13) are applicable if $t$ is assumed to be continuous for $0 \leq r \leq 1$.

[39] From the definition of $O(q^2)$ we may write

$$jHjO(j^2) = \tfrac{1}{4}\int_0^1\int_0^1 \left(\frac{1}{r_m} - 1\right) j^2(r_1)\, j^2(r_2)\, r_1{}^3 r_2{}^3\, dr_1\, dr_2$$

where $r_m = \max(r_1, r_2)$.

## REFERENCES


[1]. M. Biot, "Non-linear Theory of Elasticity," *Phil. Mag.*, Ser. 7, vol. 27 (1939), pp. 468-489; "Elastizitätstheorie zweiter Ordnung," *Z. f. ang. Math. u. Mech.*, vol. 20 (1940), pp. 89-99.

[2]. D. G. Bourgin, "The Clamped Square Sheet," *American Journal of Mathematics*, vol. 61 (1939), pp. 417-439.

[3]. R. Courant und D. Hilbert, *Methoden der mathematischen Physik*, vol. 1, 2nd ed. (1931), vol. 2 (1938).

[4]. A. Föppl, *Vorlesungen über technische Mechanik*, vol. 5, § 24 (1907).

[5]. K. O. Friedrichs and J. J. Stoker, "The Non-linear Boundary Value Problem of the Buckled Plate," *Proceedings of the National Academy of Sciences*, vol. 25 (1939), pp. 535-540.

[6]. ——— "Buckling of the Circular Plate beyond the Critical Thrust." To appear in the *Journal of Applied Mechanics*.

[7]. K. O. Friedrichs, "Über die ausgezeichnete Randbedingung in der Spektraltheorie," *Mathematische Annalen*, vol. 112 (1936), pp. 1-23.

[8]. ——— "On Differential Operators in Hilbert Spaces," *American Journal of Mathematics*, vol. 61 (1939), pp. 523-544.

[9]. A. Grzedzielski and W. Billewicz, "Sur la rigidité de la tôle flambée," *Sprawozdania Inst. Techn. Lotnictwa*, vol. 10 (1937), pp. 5-22.

[10]. A. Hammerstein, "Nichtlineare Integralgleichungen," *Acta math.*, vol. 54 (1930), pp. 117-176.

[11]. H. Hencky, "Über den Spannungszustand in kreisrunden Platten," *Z. f. Math. u. Phys.*, vol. 63 (1915), pp. 311-317.

[12]. ——— "Die Berechnung dünner rechteckiger Platten," *Z. f. ang. Math. u. Mech.*, vol. 1 (1921), pp. 81-89, 423-424.

[13]. Th. v. Kármán, "Festigkeitsproblem im Maschinenbau," *Enz. d. math. Wiss.*, Bd. IV⁴ (1910), pp. 438-352.

[14]. L. Lichtenstein, *Vorlesungen über einige Klassen nichtlinearer Integralgleichungen* (1931).

[15]. K. Marguerre und A. Kromm, "Verhalten eines Plattenstreifens oberhalb der Beulgrenze," *Luftfahrtf.*, vol. 14 (1937).

[16]. F. D. Murnaghan, "Finite Deformations of an Elastic Solid," *American Journal of Mathematics*, vol. 59 (1937), pp. 235-260.

[17]. A. Nádai, *Elastische Platten* (1925).

[18]. P. Polubarinova-Kotschina, "Zum Problem der Plattenstabilität," *Appl. Math. a. Mech.*, vol. 3 (1936).

[19]. L. Prandtl, "The Mechanics of Viscous Fluids," in Durand, *Aerodynamic Theory*, vol. 3, Sections 13, 14 (1935).

[20]. S. Timoshenko, *Theory of Elastic Stability* (1936).

[21]. E. Trefftz, "Über die Ableitung der Stabilitätskriterien," *Third International Congress of Mechanics*, Stockholm (1930).

[22]. S. Way, "Bending of Circular Plates," *Transactions of the American Society of Mechanical Engineers*, vol. 56 (1934).